Management Decision

Volume 57 Number 8 2019

Management Decision

Number 8

Big data analytics, dynamic capabilities and firm performance

Guest Editors: Sanjay Kumar Singh and Manlio Del Giudice

1729 Guest editorial

1734 Organizational capabilities that enable big data and predictive analytics diffusion and organizational performance: a resource-based perspective
Dheva Mitra, Zongwei Luo, Benjamin Hazen, Sthali Hasson and Cyril Foropon

1756 Turning information quality into firm performance in the big data economy
Samuel Fosso Wamba, Shamir Aker, Laura Trinchera and Marc De Boumont

1784 Identifying Industry 4.0 IoT enablers by integrated PCA(ISM)-DEMATEL approach
Uthmane Repeul and Surya Prakash Singh

1818 Value creation through big data in emerging economies: the role of resource orchestration and entrepreneurial orientation
Jing Zheng and Zeeran Khan

1839 Influence of basic research investment on corporate performance: exploring the moderating effect of human capital structure
Main Son, Xiongfeng Pan, Xianyou Pan and Zhiming Jiao

1857 Role of cloud ERP and big data on firm performance: a dynamic capability view theory perspective
Shreev Goyal, Shayan Dafin, Bhave Bhushan and Zongwei Luo

1880 Interplay between information systems and environmental management in ISO 14001-certified companies: implications for future research on big data
Paul De Camaergo-Frioni, Chetveli, Jose Chapiettu, Jablour, Ana Beatriz Lopez de Souza Jablour, Nelson Oliveira Stoiberi, and Emiliano Ferro

1902 Transforming big data into knowledge: the role of knowledge management practice
Roberto Chierici, Alex Holzschuch, Alexei Garcia-Perez and Siemens Worx

1923 Big data analytics capabilities and knowledge management: impact on firm performance
Abhishek Verma, Rohit Viswanath, Aseem Dhar and Jerome Coutier

1937 Big data visualization, geographic information systems and decision making in healthcare management
Anita Chinnasamy, Armando Pinto, Luca Delr and Alberto Matass

1960 The integration between knowledge management and dynamic capabilities in agile organizations
Paolo Israel, Massimo Bentivegna, Giuseppe Cossu, Riccardo Ritsema and Mirco De Pascale

1980 Big data for business management: current level and future trends
Sanjay Kumar Singh, Janger Singh and Bhalde

1995 Big data analytics and health information systems: the role of data quality and management in improving healthcare quality and efficiency
Diagnosi, M. P. P., Da Silva, N. and G. M. C.

2010 Big data analytics for business and management
Lorenzo Attardo, Veronica Scuderi, Manlio Del Giudice and Antonio Messina

2032 Big data analytics and health information systems: the role of data quality and management in improving healthcare quality and efficiency
Diagnosi, M. P. P., Da Silva, N. and G. M. C.

2052 How do different types of interorganizational ties matter in technological exploration?
Yu-Shan Su and Mike Vanheeswijk
Guest editorial

Big data analytics, dynamic capabilities and firm performance

Big data has high operational and strategic potential *vis-à-vis* business value creation and is simply the next big thing in innovation (Gobble, 2013) as it creates actionable ideas for firm performance and competitive advantage (Wamba *et al.*, 2015). Big data necessitates much more than application of new analytics (El-Kassar and Singh, 2018) as firms that learn to take advantage of big data unbridge new organizational abilities and value (Davenport *et al.*, 2012). It is also pertinent to note that corporate commitment to the use of big data analytics is very important as past literature suggest that the corporate commitment affects big data assimilation through acceptance and routinization routes (Singh and El-Kassar, 2019) and that in turn to enhance sustainable performance of the firms (Coluccia *et al.*, 2019). Several colleagues also suggest that for firms to have superior performance from their employees, they need to leverage analytics over the gut instincts (Del Giudice *et al.*, 2018; Davenport *et al.*, 2010) and that calls for a data driven decision-making culture (Santoro *et al.*, 2019; Soto-Acosta *et al.*, 2018; McAfee *et al.*, 2012). Needless to mention that use of business analytics for organizational value creation is dependent upon the roles played by the organizational decision-making processes, including resource allocation processes and resource orchestration processes (Helfat *et al.*, 2009; Teece, 1997) which requires examination and deeper understanding.

The extant literature suggests several examples of big data initiatives and treating it as firm’s dynamic capabilities that help create business relevant knowledge, to add value, enhance performance and give competitive advantage to the firms over their rivals in the dynamic market, still top managers are conspicuously reluctant to regularly allocate resources to facilitate big data analytics (El-Kassar and Singh, 2018) for sustainable development of people, process, and organization (Budhwar *et al.*, 2018; Singh, 2018; Al-Ali *et al.*, 2017). Therefore, the aforementioned literature on big data for enhancing firm dynamic capabilities and performance suggest for an urgent need to examine them in much detail. It is in this context, we believe that the papers published in this special issue on “Big data analytics, dynamic capabilities, and firm performance” try to fill in the existing gaps in the literature and also to provide direction for future researches.

The first paper titled “Organizational capabilities that enable big data and predictive analytics diffusion and organizational performance: A resource-based perspective” is contextualized in Indian context. The study suggests that big data and predictive analytic (BDPA) mediates the influence of information technology deployment and human resource capabilities on organizational performance. Therefore, it suggests that BDPA plays critical role in business decisions in firms in India.

The second paper titled “Turning information quality into firm performance in big data economy” explores how information quality dynamics links the business value, the user satisfaction and the firm performance in the big data environment. This study was conducted in two countries – France, and the USA – and extends literature on big data literature, using appraisal-emotional response-coping framework, and suggests how to use information quality modeling *vis-à-vis* firm performance.

The next paper titled “Identifying industry 4.0 IoT enablers by integrated PCA-ISM-DEMATEL approach” suggests that IoT ecosystem and IoT Big Data are the critical enablers of the industry 4.0. This study used integrated approach-based hierarchical model and cause-effect relationship amongst the IoT enablers and the authors suggest them as novel initiative for the industry 4.0.
The fourth paper titled “Value creation through big data in emerging economies: the role of resource orchestration and entrepreneurial orientation” examines how managers arrange, bundle and use key assets from big data for value creation in firms the context of the emerging economies. The key findings of the study suggests that entrepreneurial orientation is essential that firms based in the emerging economies should leverage to create value through big data by bunching together and arranging key assets to enhance firm performance.

The next paper titled “Influence of basic research investment on corporate performance: Exploring the moderating effect of human capital structure” is situated in Chinese context. The findings of the study suggest that human resource is a dynamic ability and HR practice on knowledge stock influences firms’ dynamic capability and that in turn to enhance competitiveness of the firms too.

The subsequent paper is titled “Role of cloud ERP and big data on firm performance: A dynamic capability view theory perspective.” This study provides fresh insights on the role of cloud-based ERP services and BDPA on firm performance. Furthermore, this study attest to the business relevance of using cloud ERP and big data predictive analytics to enhance and sustain firm performance in the dynamic markets.

This special issue moves to the next paper titled “Interplay between information systems and environmental management in ISO 14001-certified companies: implications for future research on big data” wherein it aims to find out how information systems contributes to the evolutionary process of corporate environmental management and its implications for big data research. This study suggests a framework that identifies the support of IS for corporate environmental practices.

The succeeding paper titled “Transforming big data into knowledge: the role of knowledge management practice” investigates on how big data collected using social media influences knowledge management practices, innovation processes and business performance. The findings of the study suggest that innovation capacity directly influences customer relationship performance. Furthermore, this study suggests that big data retrieved from social media improves both knowledge management practices and innovation capacity.

The ninth paper titled “Big data analytics capabilities and knowledge management: Impact on firm performance.” The key findings of the study suggest that organizations which have developed better big data analytic capabilities – technological and managerial – significantly enhances their performance. Furthermore, this study also suggests that knowledge management orientation plays critical role in increasing the overall effect of the big data analytic capabilities.

The next paper titled “Big data visualisation, geographic information systems and decision making in healthcare management” contributes to current scholarly debate on the value of Big Data for effective healthcare management. Based on the findings of the study, the authors suggest that existing technologies for data analytics can empower decision makers and even the public with knowledge on pollution.

The subsequent paper titled “The integration between knowledge management and dynamic capabilities in agile organizations” in contexts that requires organizational agility. Based on the findings of the study, the authors proposes a model to describe the modus operandi of a startup and enables it to develop the cycles of testing, measurement and seizure of knowledge in dynamic and uncertain contexts.

The succeeding paper titled “Big data for business management in the retail sector” investigates how does firm uses big data to transform organizational practices for the express purpose of potential benefits in the retail industry. The paper suggests that big data deployment influences business functions as the need for skilled human resources arises along with the need for data infrastructure.
This special issue moves on to another paper titled “A bibliometric analysis of research on big data analytics for business and management” examines thoroughly to classify literature linking big data analytics and management phenomena. The paper presents an interpretive framework that analyzes definitional perspectives and applications of big data analytics in the management field. The paper also suggests a general taxonomy that enhances understanding of big data vis-à-vis business value.

The 14th paper titled “Talent management under a big data induced revolution: The double-edged sword effects of challenge stressors on creativity” examines the effect of challenge stressors on creativity and the boundary conditions of the relationship. This study reveals boundary conditions by investigating dispositional and contextual factors which may accentuate the positive effect while attenuating the negative effect of challenge stressors on employee creativity.

The next paper titled “Innovating through digital revolution. The role of soft skills and big data in increasing firm performance” for developing good connect between human and technological side of the organization. The findings of the study suggest that firms’ investment in big data mediates the linkage between human resources’ organizational behavior and the organizational economic performance.

The subsequent paper titled “Big data and dynamic capabilities: A bibliometric analysis and systematic literature review” attempts to identify and address the gaps in the existing literature on big data and dynamic capabilities. The findings of the study reveals four clusters, namely, big data and supply chain management, knowledge management, decision making, business process management and big data analytics.

The succeeding paper titled “Combining organizational change management and organizational ambidexterity using data transformation” investigates and suggests for practicable data-driven theory for the implementation and management of organizational change by combining together both organization ambidexterity and change management research. The study suggests for well-grounded change management program as key tool attain competitive differentiation through ambidexterity.

The 18th paper titled “Big data analytics capability in supply chain agility: The moderating effect of organizational flexibility.” It is grounded in the dynamic capability theory and the contingency theory and the findings of the study suggest that big data analytics capability influences both supply chain agility and competitive advantage. Moreover, the results of the study make useful contributions to literature in the domain of big data analytics capability, supply chain agility and competitive advantage.

The next paper titled “Big data management: Implications of dynamic capabilities and data incubator” advances understanding on how to strategically deal with contemporary challenges of big data management, related to data veracity and data value. The authors used inductive-constructivist approach to propose a framework of data incubator to develop insights on data veracity and value.

The penultimate paper titled “Challenges with Big Data Analytics in Service Supply Chains in the UAE” examines key challenges associated with big data analytics in service supply chains in the context of the United Arab Emirates. The paper finds factors, calculate their influence over service supply chain, and suggests for adopting broad and all-inclusive view while adopting big data.

The last paper titled “How do different types of interorganizational ties matter in technological exploration?” argues for making a distinction in boundary-spanning exploration between explorative learning from partners and explorative learning from non-partners. The paper suggests that clique-spanning ties positively influences explorative learning from partners, but not on explorative learning from non-partners. Overall, this paper advances the exiting literature and suggests for future researches.
We hope that the special issue on “Big data analytics, dynamic capabilities, and firm performance” provides varied and deep insights and throw agenda for future researches.

Sanjay Kumar Singh
College of Business, Abu Dhabi University, Abu Dhabi, United Arab Emirates, and
Manlio Del Giudice
Faculty of International Business Administration, University of Rome “Link Campus”, Rome, Italy and Department of Management, Paris School of Business, Paris, France

References


About the Guest Editors
Sanjay Kumar Singh, PhD is Associate Professor of Management, College of Business, Abu Dhabi University, Abu Dhabi, UAE. His research interests include big data analytics, international and strategic HRM, knowledge management, innovation management, sustainability and ethics. Sanjay Kumar Singh is the corresponding editor and can be contacted at: sanjay.singh@adu.ac.ae

Manlio Del Giudice, PhD is Professor of Management, University of Rome “Link Campus” in Italy. He is affiliated as Research Full Professor of Entrepreneurship and Management at the Paris School of Business, in Paris, France. His research interests include knowledge management, entrepreneurship, technology transfer, cross-cultural management.
Organizational capabilities that enable big data and predictive analytics diffusion and organizational performance
A resource-based perspective

Deepa Mishra
DeGroote School of Business, McMaster University, Hamilton, Canada

Zongwei Luo
Department of Computer Science and Engineering, SUSTech, Shenzhen, China

Benjamin Hazen
Department of Marketing and SCM, University of Tennessee, Knoxville, Tennessee, USA

Elkafi Hassini
DeGroote School of Business, McMaster University, Hamilton, Canada, and

Cyril Foropon
Department of Management, Montpellier Business School, Montpellier, France

Abstract
Purpose – Big data and predictive analytics (BDPA) has received great attention in terms of its role in making business decisions. However, current knowledge on BDPA regarding how it might link organizational capabilities and organizational performance (OP) remains unclear. Drawing from the resource-based view, the purpose of this paper is to propose a model to examine how information technology (IT) deployment (i.e. strategic IT flexibility, business–BDPA partnership and business–BDPA alignment) and HR capabilities affect OP through BDPA.

Design/methodology/approach – To test the proposed hypotheses, structural equation modeling is applied on survey data collected from 159 Indian firms.

Findings – The results show that BDPA diffusion mediates the influence of IT deployment and HR capabilities on OP. In addition, there is a direct effect of IT deployment and HR capabilities on BDPA diffusion, which also has a direct relationship with OP.

Originality/value – Through this study, authors demonstrate that IT deployment and HR capabilities have an indirect impact on OP through BDPA diffusion.

Keywords Big data, Resource-based theory, Organization performance, Predictive analytics, HR capabilities, IT deployment capabilities

Paper type Research paper

1. Introduction
The topic of big data has gained immense popularity among scholars and practitioners as it has the ability to transform the entire business processes. A firm can generate business insights, enhance performance and outperform its competitors by extracting, managing and analyzing 5V data-related dimensions (i.e. volume, variety, velocity, veracity and value)
The analytical techniques that are used to gain insights from large and complex data sets in order to make better decisions are referred to as big data and predictive analytics (BDPA) (Chen et al., 2012; Watson, 2014; Frisk and Bannister, 2017). It is an increasingly important topic in today's business discourse. The decision-making capability of BDPA stimulates firm's interest in adopting data-driven decision making and advanced big data applications (McAfee et al., 2012). Certainly, vast amount of data, as well as powerful data mining tools, are available to managers and analysts (Sharma et al., 2010; Davenport et al., 2010). However, to generate insights, the mere application of analytical tools to data is not sufficient. It requires intense collaboration between analysts and managers exploiting data and analytic tools to gain more insight into it.

Although data analytics has applications in varied fields, it has been extensively applied in business. As pioneers of BDPA adoption, Google and Amazon outperformed their competitors with their ability to exploit data. Other early adopters include technology leaders such as: IBM and Hewlett-Packard (Barton and Court, 2012). More recently, the US-based BSA Software Alliance reported over 10 percent increase in growth for 56 percent of firms that implemented BDPA (Columbus, 2014). Moreover, there are various evidence on the adoptions and implementations of BDPA in different industries. For instance, Wixom et al. (2013) studied its successful application in a fashion industry and Anderson-Lehman et al. (2004) described its use in airlines.

Due to its well-perceived capability and popularity in the business world, there has been a rapid growth of interest in BDPA in the academic community. That interest is often centered on the effects of BDPA on business process, workforce effectiveness, cost savings, competitive advantage and performance (Wang et al., 2016). In particular, BDPA helps organizations in understanding its business and markets and also in extracting maximum opportunities offered by numerous data and domain-specific analytics (Chen et al., 2012, pp. 1166–1168). While some studies have suggested the significance of BDPA in enhancing organizational performance (OP), improving decision making and transforming the supply chain (Schoenherr and Speier-Pero, 2015; Bose, 2009; Waller and Fawcett, 2013), others have provided evidence that leading firms outdo their peers by identifying different ways of leveraging BDPA (McGuire et al., 2012). These tools can offer better decision making, uncover valuable information, identify customer needs and preferences, and develop advanced products.

Numerous articles explore the impact of organizational capabilities on competitive advantage and OP (Wu et al., 2006; Rungtusanatham et al., 2003; Ravichandran and Lertwongsatien, 2005; Kim et al., 2011; Xu et al., 2014). For instance, firms can sustain or achieve competitive advantage if they properly utilize their capabilities (Wu et al., 2006). Drawing from resource-based theory (RBT), Ravichandran and Lertwongsatien (2005) described the manner in which information systems resources and capabilities can improve OP. In the same vein, Kim et al. (2011) empirically tested the relationships among information technology (IT) capabilities, process-oriented dynamic capabilities, and financial performance and hence investigated the strategic role of IT in enhancing OP. Recently, Xu et al. (2014) studied the effect of top management support and IT on leveraging supply chain integration and business performance.

Although it seems reasonable to expect a relationship between IT capabilities and performance, the role of BDPA diffusion in this relationship is not clear. In this paper, we are taking the first step in this direction by analyzing these relationships. We empirically analyze three key questions: Do organizational capabilities (IT deployment and human resource (HR)) enable diffusion of BDPA? What role might BDPA diffusion play in regard to the relationship between organizational capabilities and performance? What is the impact of BDPA diffusion on OP? To answer these questions, we construct a conceptual model and
test it using multiple regression analysis, using data collected from 160 Indian companies. The results contribute to the BDPA literature by investigating to what extent capabilities of an organization affect BDPA diffusion and the impact of this diffusion on OP.

The rest of this paper is organized as follows. First, the authors use theory and literature to develop the conceptual model, which integrates IT deployment and HR capabilities, BDPA diffusion and OP. Then, hypotheses are further developed. Next comes a discussion of the research methodology including the data collection procedure. After reporting the results, the paper ends with a brief discussion section, which includes theoretical and practical implications, as well as limitations and some concluding remarks.

2. Literature review

2.1 Organizational capabilities

The RBT is one of the most commonly used theories for understanding how firms achieve and sustain their competitive advantage (Barney et al., 2011). It conceptualizes firms as a bundle of resources and capabilities that are distributed in a heterogeneous manner across the firms and these differences continue to hold over the time (Barney, 2001; Ambrosini and Bowman, 2009). As such, firms having valuable, rare, inimitable and non-substitutable resources and capabilities can achieve sustainable competitive advantage by applying value-creating strategies (Eisenhardt and Martin, 2000; Peteraf and Barney, 2003; Cheng and Shi, 2015).

Scholars (Cepeda and Vera, 2007; Zahra et al., 2006) have classified organizational capabilities into two types; dynamic and operational. The dynamic capability refers to the ability of a firm to achieve new resource conditions with the changing market scenarios (Martelo Landroguez et al., 2011). Organizations focus on developing dynamic capabilities in order to gain a sustained competitive advantage (Eisenhardt and Martin, 2000; Cheng et al., 2014). In fact, dynamic capabilities extend resource-based view (RBV) since it is inadequate to explain how and why certain organizations gain the competitive advantage when the markets are dynamic and cannot be predicted (Eisenhardt and Martin, 2000). On the other hand, operational capability represents how firms operate to make a living in the present (Wu et al., 2010). It represents the ability of a firm to accomplish and coordinate the tasks that are necessary to carry out operational activities; distribution logistics and marketing campaigns (Pavlou and El Sawy, 2006). Since the necessity for delivering cost-effective products on time is increasing, organizations are building operational capabilities to improve its performance (Ngai et al., 2011; Overby et al., 2006).

2.2 IT and HR Capabilities

IT capabilities are defined as the high performing organizational processes that acquire, deploy and leverage IT assets, such as technical and human assets (Pavlou and El Sawy, 2006; Bingham et al. 2007; Tian et al., 2010). Prior research works (Carr, 2003; Sook-Ling et al., 2015) reveal that IT assets which are easy to imitate cannot improve the firm’s competitive advantage as these resources can easily be copied by the competitor firm. In this regard, Wade and Hulland (2004) suggested that even if the imitable assets contribute to improve the competitive advantage, it will not be long-lasting.

IT capabilities help a firm in supporting or shaping its business as it deals with acquiring, deploying and leveraging of IT resources (Pavlou and El Sawy, 2006), which can be acquired through outsourcing or systems development (Aubert et al., 2008; Wullenweber et al., 2008; Goo et al., 2009; Patnayakuni and Ruppel, 2010). Some authors have also discussed leveraging of IT resources and IT deployment capabilities. For instance, Tian et al. (2010) argued that IT capabilities allow a firm to “configure and reconfigure the IT infrastructure, by adding new IT components or by adapting existing information systems,
in order to make the whole collection of information system available to support and shape businesses” (p. 241). In literature, scholars (Bhatt and Grover, 2005; Kim et al., 2011; Fosso Wamba et al., 2016) have identified various typologies for IT capabilities. For instance, IT capability can be well explained via value, heterogeneity and imperfect mobility, where the first two reflects the necessary conditions for achieving competitive advantage while, the last is important for obtaining sustained advantage (Bhatt and Grover, 2005). In addition, IT capabilities are classified as value capability, competitive capability and dynamic capability (Liu et al., 2013).

Another important organizational capability is its HR which encompasses resources, relationships and decisions that allow firms to take chance to gain an advantage in the market and maintain competition (Meyer and Dunphy, 2016). Khandekar and Sharma (2005) suggested that HR plays an important role in maintaining the competitive advantage as it is the crux of competitive strategy. In view of RBT, firms have the ability to understand whether skills and knowledge of employees are appropriate or not. The importance of HR capability can also be identified from the work of Bernardin and Russell (2013) who mentioned that competitive advantage can be maintained if firms attract and retain skilled workers who provide them a competitive edge.

From the aforementioned studies, it can be clearly seen that the ability of a firm to innovate is majorly influenced by the knowledge and skills of its employees (Khandekar and Sharma, 2005). On the other hand, the less qualified employees are the biggest barriers to innovation (OECD, 2000). Therefore, HR can be seen as one of the major drivers for successful new product development (Epstein et al., 2015).

2.3 BDPA diffusion and organizational performance

BDPA has been identified as a strategic tool that has the potential to improve business efficiency and effectiveness (Fosso Wamba et al., 2017). In their study on big data, Ji-fan Ren et al. (2017) emphasized that organizations with successful implementation of BDPA can enhance their product development by 70 percent, market expansion 72 percent, customer satisfaction 79 percent and economic performance in terms of sales and revenue by 76 percent. In addition, Germann et al. (2014) noted that BDPA diffusion is positively related with the firm performance as it allows organizations to analyze and manage their strategy through the lens of data (Gunasekaran et al., 2017, 2018). In their study on IT capabilities, Kim et al. (2011) highlighted that there exists a positive relation between IT capability and firm performance; business process and financial. Recently, Fosso Wamba et al. (2016) established that IT capability, firm performance, firm agility and stock market returns are related. Thus, the importance of BDPA can be adjudged from the fact that it is now used as a differentiating factor between firms in terms of their performance.

3 Conceptual model and hypotheses development

Rooted in strategic management literature, the RBV considers a firm as a network of resources and capabilities which cannot be easily bought nor sold in the market. Although these resources and capabilities offer several benefits that cannot be imitated by the competitors, they are the potential source of competitive advantage. This theory assumes that the resources required for selecting and implementing strategies are immobile and not evenly distributed across firms. Moreover, the terms “resources” and “capabilities” are often used interchangeably in the literature (Wu et al., 2010). While resources are the building blocks of analyses, organizational capabilities build on the interaction of these resources to create competitive advantage (Bharadwaj, 2000; Cavusgil et al., 2007; Schreyögg and Kliesch-Eberl, 2007).

It is vital to leverage resources and capabilities to generate superior firm performance. Researchers, drawing largely on a resource-based perspective, have argued that superior
firm performance is obtained from two types of organizational capabilities: operational capabilities and dynamic capabilities (Eisenhardt and Martin, 2000; Helfat and Peteraf, 2003; Zahra et al., 2006; Cepeda and Vera, 2007; Essex et al., 2016). While operational capability refers to the ability of a firm to accomplish and coordinate various operational activities, a dynamic capability represents the firm's ability to combine, develop and reassemble internal and external competencies in order to deal with the rapidly changing environment (Teece et al., 1997). In the literature, BDPA is proposed as a dynamic capability (Chen et al., 2015). With this idea, it gets easy to understand the implications of BDPA diffusion on organizational value creation. Moreover, the concept of the hierarchy of capabilities posits that a higher-order capability develops from various lower-order capabilities (Grewal and Slotegraaf, 2007; Sirmon et al., 2007; Kraaijenbrink et al., 2010). In view of this discussion, we propose a conceptual model where lower-order capabilities (IT and HR) are leveraged to develop higher-order capability (BDPA) that, in turn, directly affects OP (Figure 1).

3.1 IT deployment capabilities
IT capabilities refer to the organizational processes that acquire, deploy and leverage IT assets to enhance performance (Wade and Hulland, 2004, Pavlou and El Sawy, 2006; Bingham et al., 2007). It is a major factor that not only influences a firm’s growth and survival but also enables successful firms to compete with their peers. While a stream of research has focused on acquiring IT resources using outsourcing or system development (Aubert et al., 2008; Wullenweber et al., 2008; Patnayakuni et al., 2007), the other stream focuses on leveraging IT resources. In this study, we consider IT deployment capabilities which rebuild the IT infrastructure of a firm by either introducing new IT components or by modifying current information systems. Here, we consider IT deployment capabilities to be made up of three independent components; strategic IT flexibility, business–BDPA partnership and business–BDPA alignment (Figure 2).

3.2 Strategic IT flexibility
Strategic IT flexibility is the ability of an organization to fulfill different IT demands from uncertain environments. It calculates the extent to which an organization can do IT-related activities easily and speedily to deal with the changing business, technologies and environment (Tian et al., 2010).

We propose strategic IT flexibility as an IT deployment capability where deployment refers to the process of installing and delivering IT in the concerned organization (Lai and
Mahapatra, 1997). In this vein, Tian et al. (2010) noted that new information technologies can be successfully installed and associated IT services can be easily delivered if organizations have a flexible information infrastructure. According to Ravichandran and Lertwongsatien (2005), flexibility facilitates fast and smooth application of innovative technologies in order to assist, encourage and improve ever-changing businesses. Hence, strategic IT flexibility influences companies to invest in organizational capabilities that help in responding to the uncertain environment (Kogut and Kulatilaka, 2001).

Drawing from the dynamic capabilities perspective, we propose BDPA diffusion as an organizational capability that serves to minimize uncertainties in demands, capacities and supply availability. Specifically, an organization implementing BDPA is likely to build information processing capabilities which assist them in understanding and combining knowledge obtained from different sources and directing this synthesized knowledge toward suitable decision makers (Schoenherr and Swink, 2012). Through BDPA, organizations can forecast future demands and requirements, thereby, making way for more precise and robust resource configurations (Chen et al., 2015, Wang and Wei, 2007). Since making strategic and operational decisions in an uncertain environment is a critical task for managers involved in the decision-making process (Carmeli et al., 2011), BDPA plays a major role in this scenario as it helps an organization’s managers to make better decisions (Davenport, 2006; Davenport and Harris, 2007). In addition, Eisenhardt and Martin (2000) suggest that in uncertain environments, dynamic capabilities often depend on the new knowledge created from a particular situation rather than on existing knowledge. This condition provides additional support for a potential impact of strategic IT flexibility on BDPA diffusion. Therefore, we hypothesize:

H1. Strategic IT flexibility has a positive effect on BDPA diffusion.

3.3 Business–BDPA partnership

Business–BDPA partnership is an organizational capability that enables the smooth operation of the complete IT deployment process. Here, partnering refers to a collection of actions performed repeatedly to integrate the efforts of IT (in our case, BDPA) and business units for deploying IT in order to aid, encourage and improve businesses (Tian et al., 2010). We suggest business–BDPA partnership as an IT deployment capability based on Cooper and Zmud (1990), who described deployment as the efforts made by an organization to promote easy diffusion of IT within a user community. In this vein, Bhatt and Grover (2005) and Bassellier et al. (2001) pointed out that IT resources can be successfully deployed if IT
departments are interested in satisfying business needs and demands, and maintaining a strong relationship with the business department. Further, business and IT functions should be eager to work together in order to get an idea about the method through which cutting-edge information technologies can be utilized to observe and capture upcoming business opportunities. Hence, strong partnership facilitates successful implementation of IT (Bharadwaj et al., 1999).

Drawing from the recent report published by New Vintage Partners (NPV, 2016), we propose that successful diffusion of BDPA depends upon the partnership between business and technology organizations. They reported that 33.9 percent firms considered the cooperation between business and technology as the most crucial element for enabling business adoption. Next, 23.2 percent firms observed strong business sponsorships as the second important element. On the other side, factors such as technology leadership selection received negligible attention. Thus, it is clearly observable that partnership and cooperation with business leaders might be important factors to BDPA diffusion (NPV, 2016). Therefore, we hypothesize:

\[ H2. \] Business–BDPA partnership has a positive effect on BDPA diffusion.

### 3.4 Business–BDPA alignment

Business–BDPA alignment refers to the ability of an organization to assign resources and build strategies in an attempt to properly align BDPA and business (Tian et al., 2010). We suggest business–BDPA alignment as an IT deployment capability because deployment is the adoption of suitable information technologies that can aid and encourage business and supply chain activities. Although BDPA supports business, it is clear that business should also capitalize on strategic potentials of BDPA. Hence, business–BDPA alignment is a bidirectional process which can be defined at both the process level and organizational level.

Despite the widespread attention from both researchers and professionals, business and BDPA alignment pose a challenge to many organizations. This may be due to changing business needs, fluctuating market conditions, new technology and the difficulty that the organization faces while dealing with these changes. Moreover, to avail the benefits of investments in BDPA solutions, we posit that enterprises might wish to focus on bridging the gap between analytics and operational needs. To this end, it would be important for an organization to maintain a reliable database that mixes data from all other sources and helps in storing data clean, current and complete data. Hence, maximum insights from BDPA can be obtained if business, IT and data professionals work together. In this vein, Weldon (2016) notes that “business and IT alignment is a key for data analytics success.” Therefore, we hypothesize:

\[ H3. \] Business–BDPA alignment has a positive effect on BDPA diffusion.

### 3.5 Human resource (HR) capabilities

Academic literature has identified HR as a key factor for attaining competitive advantage (Barney and Clark, 2007, Kamoche and Mueller, 1998). In this vein, Newbert (2007) stated that among many organizational capabilities, HR is one of the most widely studied capabilities, and generally can be linked to competitive advantage. Much attention is currently being paid to the effect that strategic involvement of HR could create on OP. This growing interest has triggered the development of a resource-based model for HR management (Boxall, 1996), to the extent that HR has been identified as the reason for the success of an organization (Kakabadse and Kakabadse, 2000) and as an indicator of enhanced organizational effectiveness (Analoui, 1999; Analoui, 2002).
RBV draws attention to the importance of internal resources of organizations in enhancing competitive advantage. Past research also describes the role of executives in building dynamic capabilities (Kor and Mahoney, 2005; Teece, 2007; Augier and Teece, 2009), specifically in redeveloping the resource base (Ambrosini and Bowman, 2009; Moliterno and Wiersema, 2007). Certainly, a wrong decision by an executive in any situation may lead to the wrong dynamic capability which could be harmful to a firm (Adner and Helfat, 2003; Ambrosini and Bowman, 2009). Moreover, availability of skilled professionals with the ability to carry out business analytics reflects upon the willingness of organizations to diffuse BDPA (Chen et al., 2015). In order to get full support from top managers, it is desired that an organization must have enough resources and capabilities to promote the diffusion of BDPA. Therefore, we hypothesize:

\[ H4. \] HR capabilities have a positive effect on BDPA diffusion.

### 3.6 Impact of capabilities on organizational performance

From an RBV perspective, lower-order capabilities help the firm in building higher-order capabilities that, in turn, enhance OP (Liu et al., 2013). For instance, IT capabilities considered as lower-order capabilities leverage development of high-order capabilities which helps the firm in achieving high performance (Rai et al., 2006). Further, Liang et al. (2010) studied the impact of IT and organizational resources on OP. Moreover, literature has also acknowledged that organizational capabilities mediate the relation between IT resources and OP (Collis, 1994). In their study on capabilities, Wu et al. (2006) mentioned that supply chain capabilities mediate the relation between the IT-related resource and firm performance. Hence, in this study, we suggest that lower-order capabilities (IT deployment and HR) might lead to OP when mediated by BDPA diffusion (higher-order capability). Therefore, we hypothesize:

\[ H5a. \] Strategic IT flexibility under the mediation effect of BDPA diffusion is positively related to OP.

\[ H5b. \] Business–BDPA partnership under the mediation effect of BDPA diffusion is positively related to OP.

\[ H5c. \] Business–BDPA alignment under the mediation effect of BDPA diffusion is positively related to OP.

\[ H5d. \] HR capabilities under the mediation effect of BDPA diffusion is positively related to OP.

### 3.7 BDPA diffusion and organizational performance

The extant resource-based research identified resource complementarity as a major factor for enhancing business value and OP (Morgan et al., 2009; Ji-fan Ren et al., 2017). Past literature also suggests that IT resources help in improving business value and influence the OP (Grant, 1991; Barua et al. 1995; Mooney et al., 1996; Ji-fan Ren et al., 2017). Also, these resources can be major drivers for improving efficiency and effectiveness, which further lead to the OP (Chang and King 2005; Jayachandran et al., 2005). The impact of IT capabilities on OP was examined by Liu et al. (2013), who viewed the problem through the lens of organizational capabilities. Hence, OP is the central construct in RBT (Kozlenkova et al., 2014). Due to these reasons, the relevance of RBV is significant in the BDPA environment. Moreover, OP can be fostered by the technology resources in BDPA (Kiron et al., 2014; McAfee et al., 2012). Therefore, we hypothesize:

\[ H6. \] BDPA diffusion has a positive impact on OP.
4. Research methodology

We collected data from Indian firms which were later analyzed in order to test our proposed hypotheses. In this section, we provide the description of the survey instrument and data collection plan.

4.1 Survey instrument design

For conducting this study, we divided the survey instrument into three broad sections. In the first section, we covered the constructs related to organizational capabilities (IT deployment and HR), and in the second and third sections, we included the constructs BDPA diffusion and OP, respectively. The items were assessed by participants based on Likert scales ranging from “strongly disagree” to “strongly agree.”

Each construct was modeled using multi-item reflective scales (DeVellis, 2003), which were developed after conducting a thorough review of existing scales and a pre-test study involving ten experts to make sure that the involved constructs were valid and functional. Next, the pilot study was conducted by randomly selecting 60 respondents from the sample frame to ensure the reliability of the survey instrument. All the items used in the survey instrument are presented in Table I.

4.1.1 IT deployment capabilities. In this research, we measured deployment capabilities through strategic IT flexibility, business–BDPA partnership and business–BDPA alignment. We developed four items to represent strategic IT flexibility, four items for business–BDPA partnership and six items for business–BDPA alignment based upon research by Tian et al. (2010).

4.1.2 HR capabilities. The construct “HR capabilities” was measured using four items focusing on knowledge and skill of employees, and their commitment toward the organization, based on Wiklund and Shepherd (2003).

4.1.3 BDPA diffusion. This construct was assessed using five items which examine the degree of BDPA diffusion within accounting management, product and service delivery management, warehousing and inventory management, production and operations management, and purchase and fulfillment management, based on research by Hazen et al. (2012).

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Definitions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic IT flexibility</td>
<td>The organizational capability which aids in modifying information systems according to environmental changes by either combining new IT components with the existing IT infrastructure or by re-building the existing information systems</td>
<td>Tian et al. (2010)</td>
</tr>
<tr>
<td>Business–IT partnership</td>
<td>The organizational capability considered as an organizational social capital which helps in smoothing the complete IT deployment process</td>
<td>Tian et al. (2010)</td>
</tr>
<tr>
<td>Business–IT alignment</td>
<td>The organizational capability to assign resources and build strategies to attain and maintain alignment between IT and business so that it can support and shape business</td>
<td>Tian et al. (2010)</td>
</tr>
<tr>
<td>Human resource capability</td>
<td>Describes the role of executives in building dynamic capabilities, specifically in redeveloping the resource base</td>
<td>Wiklund and Shepherd (2003), Teece (2007) and Augier and Teece (2009)</td>
</tr>
<tr>
<td>BDPA diffusion</td>
<td>Describes the levels of adoption and use of the analytical techniques that are used to gain insights from large and complex data sets to make better decisions</td>
<td>Hazen et al. (2012), Chen et al. (2012) and Watson (2014)</td>
</tr>
</tbody>
</table>

Table I. Constructs
4.1.4 Organizational performance. It was measured using three items: average market share, average sales volume and average sales growth, that reflect the extent to which focal firm performs better than its competitors (Whitten et al., 2012).

4.1.5 Control variables. In this research, we considered the firm type and size as the two important control variables (Liang et al., 2007) and took their log for our analysis purpose (Liu et al., 2013). Firms are different from each other with respect to their management and strategies so we viewed them based on the SIC code. The firm size may be crucial, so we measured it using the number of employees and annual revenue of the firm (Zhou and Li, 2010).

4.2 Data collection
We collected data by sending an invitation letter to 383 respondents working in Indian firms was obtained from databases provided by Confederation of Indian Industry and Dun & Bradstreet. In all communication, we assured our respondents to maintain strict anonymity and confidentiality. The invitation letter and the explanation at the beginning of the survey stated that in order to participate in the study, respondent’s organization should have adopted BDPA. Upon identifying respondents as potential key informants, we qualified them by analyzing how knowledgeable they deem themselves about their own firm and their level of involvement in the adoption of BDPA within their firm. Our survey mainly targeted the manufacturing, consulting, e-commerce and technology companies. In this way, we aim to ensure the validity of our findings while maintaining the homogeneity of the working environment for the respondents. Data were collected following a modified version of Dillman’s (2007) total design method. Overall, we obtained 160 usable responses resulting in a 41.67 percent response rate, which is generally typical of survey-based studies (Olson et al. 2005; Cannon and Perreault, 1999). The survey targeted experienced personnel, such as functional heads associated with supply chain management and operations management divisions. Table II shows the detailed demographics of the respondent firms.

4.3 Non-response bias treatment
To check for the potential non-response bias, we conducted a successive wave analysis in which the answers of the early respondents are compared with the answers of late respondents of the survey, where the late respondents are treated as non-respondents (Armstrong and Overton, 1977). We found that there was no significant difference ($p > 0.10$) between answers of early and late respondents, indicating that non-response bias is not a matter of concern for our research (Everaert et al., 2007; Baihaqi and Sohal, 2013).

5. Data analysis and results
5.1 Analysis of psychometric properties
Before checking the validity and reliability of the survey instrument, it is necessary to check the assumptions of homogeneity of variance, the presence of outliers and normality. To test for the homogeneity of variance and univariate normality of the constructs, we used residual plots and calculated statistics for skewness and kurtosis. Further, we detected the outliers using Mahalanobis distance of predicted variable (Cohen et al., 2003). We observed the statistics value as $0.657 < 2$ and $1.252 < 7$ corresponding to univariate skewness and kurtosis. Thus, we may conclude that univariate non-normality was not a major problem in our case (Curran et al., 1996; Dubey and Gunasekaran, 2015). Moreover, the residual plots depict that there is no significant deviation from the assumption of homoscedasticity. We calculated Mardia’s (1970) coefficient for testing multivariate normality and found that $p < 0.05$ suggesting that the value is non-significant (Blome et al., 2013). Next, to test
multicollinearity, we calculated the variance inflation factors (VIF) for each construct and observed that the values ranged from 1.000 to 7.974 (< 10), indicating that multicollinearity did not pose a problem (Hair et al., 1995).

We used factor analysis to test construct reliability, convergent validity, unidimensionality and discriminant validity. For evaluating internal consistency, convergent validity and discriminant validity, we compute Cronbach’s $\alpha$, scale composite reliability (SCR) and average variance extracted (AVE) for each construct. Standardized loadings and reliabilities of the constructs in the model are shown in Table III. We find that factor loadings of all the items are greater than 0.50, Cronbach’s $\alpha$ and SCR are greater than 0.70 (Götz et al., 2010), and AVE for each construct is more than 0.50. Therefore, the notion of construct reliability and convergent validity can be considered satisfied (Aloini et al., 2011). Also, unidimensionality was supported as AVEs are greater than 0.50 and composite reliabilities are greater than 0.70 (Ji-fan Ren et al., 2017). We further checked discriminant validity (Table IV) by calculating the square root of AVE of each construct (italic values in the leading diagonal) and the correlation coefficients between two constructs (values below the leading diagonal). As the values in the leading diagonal are greater than all the values in the same column and row, we can say that discriminant validity is possessed by our constructs (Fornell and Larcker, 1981).

### 5.2 Hypotheses testing

In order to test our proposed hypotheses, we used multiple regression along with the mediation test (Table V). We adopted this multiple regression analysis as it is the most suitable technique while dealing with complex models and available data points (Dubey et al., 2015). For testing multicollinearity, we computed VIF for each coefficient and the values were found to range from 1.000 to 7.947, which are significantly below the threshold value of 10 (Hair et al., 2006).
Furthermore, we test our hypotheses using hierarchical mediation regression analyses by following the steps of Baron and Kenny (1986).

H1 suggests that strategic IT flexibility is positively associated with BDPA diffusion. Since the value of path coefficient is 0.446 (p < 0.001), we find support to H1. H2 suggests that business–BDPA partnership has a positive association with BDPA diffusion. As the value of path coefficient is 0.672 (p < 0.001), we find support to H2.
H3 argues that business–BDPA alignment is positively associated with BDPA diffusion. As the value of path coefficient is 0.809 ($p < 0.001$), we find support to H3. H4 suggests that HR capabilities have a positive association with BDPA diffusion. As the value of path coefficient is 0.935 ($p < 0.001$), we find support to H4.

For H5a, we first performed regression analysis with strategic IT flexibility as an independent variable and OP as a dependent variable (path C) and found that strategic IT flexibility has a significant influence on OP ($\beta = 0.467; p < 0.001$). The next step was to test the effect of strategic IT flexibility on BDPA diffusion (path A), which showed significant influence with $\beta = 0.446$ ($p < 0.001$). In the third step, the influence of BDPA diffusion on OP (path B) was tested and the result was found to be significant with $\beta = 0.281$ ($p < 0.001$). Finally, the last step tested the effect of strategic IT flexibility on OP by controlling BDPA (path C'), which resulted in $\beta = 0.090$ ($p > 0.1$). It can be observed that the relations in the first three steps are significant while the relation in the last step is not significant. Thus, we find support to H5a since BDPA diffusion acts as a full mediator between strategic IT flexibility and OP (see Table V).

For H5b, we follow the steps of H5a and find that in the first three steps, the relations are significant while it is not significant in the last step $0.028$ ($p > 0.1$). Hence, we find support to H5b since BDPA diffusion acts as a full mediator between business–BDPA partnership and OP (see Table V). For H5c, we follow the steps of H5a and find that in all the four steps, the relations are significant, hence indicating partial mediation. Hence, we find support to H5c (see Table V).

<table>
<thead>
<tr>
<th>Constructs</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Strategic IT flexibility</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Business-IT partnership</td>
<td>0.28</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Business–IT alignment</td>
<td>−0.12</td>
<td>−0.07</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. HR capability</td>
<td>0.46</td>
<td>0.34</td>
<td>0.00</td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. BDPA diffusion</td>
<td>−0.12</td>
<td>−0.11</td>
<td>0.39</td>
<td>0.08</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>6. Organizational performance</td>
<td>0.17</td>
<td>0.06</td>
<td>−0.11</td>
<td>−0.13</td>
<td>−0.07</td>
<td>0.68</td>
</tr>
</tbody>
</table>

**Table IV.** Discriminant validity matrix

<table>
<thead>
<tr>
<th>Path</th>
<th>$R$</th>
<th>$R^2$</th>
<th>$\beta$</th>
<th>$p$</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic IT flexibility $\rightarrow$ OP</td>
<td>0.467</td>
<td>0.218</td>
<td>0.467</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td>Strategic IT flexibility $\rightarrow$ BDPA</td>
<td>0.446</td>
<td>0.199</td>
<td>0.446</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td>BDPA $\rightarrow$ OP</td>
<td>0.281</td>
<td>0.079</td>
<td>0.281</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td>(Strategic IT flexibility + BDPA) $\rightarrow$ OP</td>
<td>0.474</td>
<td>0.225</td>
<td>0.090</td>
<td>0.255</td>
<td>1.249</td>
</tr>
<tr>
<td>Business–BDPA partnership $\rightarrow$ OP</td>
<td>0.394</td>
<td>0.156</td>
<td>0.394</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td>Business–BDPA partnership $\rightarrow$ BDPA</td>
<td>0.672</td>
<td>0.452</td>
<td>0.672</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td>BDPA OP</td>
<td>0.281</td>
<td>0.079</td>
<td>0.281</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td>(Business–BDPA partnership + BDPA) OP</td>
<td>0.395</td>
<td>0.156</td>
<td>0.028</td>
<td>0.778</td>
<td>1.825</td>
</tr>
<tr>
<td>Business–BDPA alignment OP</td>
<td>0.167</td>
<td>0.028</td>
<td>0.167</td>
<td>0.034</td>
<td>1.00</td>
</tr>
<tr>
<td>Business–BDPA alignment $\rightarrow$ BDPA</td>
<td>0.809</td>
<td>0.655</td>
<td>0.809</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td>BDPA OP</td>
<td>0.281</td>
<td>0.079</td>
<td>0.281</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td>(Business–BDPA alignment + BDPA) OP</td>
<td>0.299</td>
<td>0.089</td>
<td>0.422</td>
<td>0.001</td>
<td>2.892</td>
</tr>
<tr>
<td>HR capability $\rightarrow$ OP</td>
<td>0.224</td>
<td>0.050</td>
<td>0.224</td>
<td>0.005</td>
<td>1.00</td>
</tr>
<tr>
<td>HR capability $\rightarrow$ BDPA</td>
<td>0.935</td>
<td>0.875</td>
<td>0.935</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td>BDPA $\rightarrow$ OP</td>
<td>0.281</td>
<td>0.079</td>
<td>0.281</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td>(HR capability + BDPA) $\rightarrow$ OP</td>
<td>0.301</td>
<td>0.091</td>
<td>0.569</td>
<td>0.009</td>
<td>7.974</td>
</tr>
</tbody>
</table>

**Table V.** Results of mediation test
For $H_{5d}$, we follow the steps of $H_{5a}$ and find that in all the four steps, the relations are significant. Hence, we find support to $H_{5d}$ since BDPA diffusion acts as a partial mediator between HR capabilities and OP (see Table V).

Finally, $H_{6}$ argues that BDPA diffusion has a positive impact on OP. Since the value of path coefficient is $0.281 (p < 0.001)$, we find support to $H_{6}$, which is consistent with the findings of Ji-fan Ren et al. (2017).

6. Discussion
Drawing on the theoretical insights of the RBV, in this paper, we have attempted to show how IT deployment and HR capabilities may contribute to OP by facilitating the diffusion of BDPA. Our results provide strong support for our first hypothesis which suggests that strategic IT flexibility effects BDPA diffusion. We find that strategic IT flexibility is positively associated with BDPA diffusion which is consistent with previous studies showing that the BDPA diffusion plays an important role in minimizing uncertainties in demands, capacities and supply availability (Chen et al., 2015; Wang and Wei, 2007). Further, BDPA helps organizations in making better decisions by predicting future demands and needs in an uncertain environment (Davenport, 2006; Davenport and Harris, 2007). The positive association between business–BDPA partnership and BDPA diffusion is consistent with the assumption that partnership and cooperation with business leaders are the keys to big data success (NPV, 2016). We further find that business–BDPA alignment and BDPA diffusion are positively associated which is in support to the suggestion that maximum insights from BDPA can be obtained when business, IT and data professionals work together (Weldon, 2016). We also find support for $H_{4}$. An important implication here is that the availability of skilled professionals with the ability to carry out business analytics reflects upon the organizational readiness to diffuse BDPA (Chen et al., 2015). In addition, through this study, we have also understood the conceptual difference between business–BDPA partnership and business–BDPA alignment. Business–BDPA partnership is the cooperation between two companies where they both utilize resources and technologies for the successful joint implementation of BDPA. In that way, both companies derive benefits from BDPA. On the other hand, through business–BDPA alignment, we refer to how firms can strategically arrange their strategies, capabilities, resources and management systems internally so that they can successfully implement BDPA within the company.

Our results provide support for the fifth set of hypotheses. With reference to $H_{5a}$ and $H_{5b}$, BDPA diffusion fully mediates the relation between IT deployment capabilities and OP. However, it partially mediates the relation between business–BDPA alignment and OP, as well as, between HR capabilities and OP. Furthermore, our prediction that BDPA diffusion enhances OP is supported: BDPA is positively associated with OP. A possible justification is that it might not be just the diffusion, but actually the proper and informed utilization of BDPA that leads to higher levels of OP. Ji-fan Ren et al. (2017), for example, posited that there is a positive impact of BDPA usage on firm performance. Consistent with this explanation, Bose (2009) and Schoenherr and Speer-Pero (2015) noted that BDPA plays an important role in making high-quality decisions and improving performance. Although here we find that diffusion of BDPA is related to performance, the next logical step for research will be to examine precisely how and why this is the case.

6.1 Theoretical contributions and practical implications
Our study provides four major contributions to big data community. First, it bridges often disparate discussions on IT deployment capabilities, HR capabilities, BDPA diffusion and OP. The findings reflect that while these two capabilities do not directly affect OP, they do so indirectly through BDPA diffusion. Second, this study enriches our understanding of the relation between IT deployment and HR capabilities and BDPA diffusion. Consistent with
prior studies (Liu et al., 2013), our findings confirm that lower-order capabilities (IT deployment and HR) help in building higher-order capabilities (BDPA diffusion). Third, this study also highlights the importance of RBV in understanding the concept of BDPA diffusion and organizational capabilities (Barney, 2014), and investigates how BDPA diffusion affects OP directly. The extant literature has examined effects of BDPA on firm performance (see e.g. Bose, 2009; Schoenherr and Speier-Pero, 2015; Ji-fan Ren et al., 2017). Through our research, we provide empirical justification to the notion that higher-order dynamic capabilities may directly affect OP. Last, we draw on the literature highlighting the role of lower and higher-order capabilities for achieving OP (Liu et al., 2013).

From a practical point of view, our study provides guidance to managers involved in the BDPA implementation process. The mediating role of BDPA diffusion clearly highlights how it can be leveraged as a source of OP. The finding that the three IT deployment capabilities and HR capabilities strongly influence BDPA diffusion; thereby improving the firm performance, indicates that managers need to concentrate on organizational capabilities. Moreover, our finding that IT deployment and HR capabilities positively affect BDPA diffusion is beneficial for firms that invest heavily in IT and HR to achieve superior OP as these investments may not be fruitful if organizations do not leverage their capabilities to attain superior dynamic capabilities. Thus, it is crucial for managers to leverage lower-order capabilities (IT and HR) to build higher-order organizational capabilities (BDPA diffusion), and improve OP. Also, this research provides insights to managers on how BDPA diffusion can directly influence OP. Although potential benefits of BDPA diffusion are well recognized in industries, there are some who are reluctant to use it due to their insufficient knowledge about the way to proceed ahead and implement BDPA. Our finding provides the necessary guidance and assurance that BDPA usage can benefit the organization.

6.2 Limitations and future research directions
While this study is underpinned by a sound theory and research methods, it has some limitations. First, cross-sectional data were used to test our research hypotheses because we were not able to collect longitudinal data. In that case, future research can include a temporal distance between data collections. Second, although we used a robust sampling procedure, caution should be taken to generalize findings beyond the research context. Therefore, we believe that researchers can expand this study by collecting data from both developed and developing countries and by comparing it with our findings. Third, we have focused only on two types of organizational capabilities—IT deployment and HR. However, one could expand this to other types of capabilities as well. Fourth, in the present study, we have considered only OP. Scholars can expand further to examine the impact on supply chain performance. Lastly, scholars could also explore the relationship between business–BDPA partnership and business–BDPA alignment as they are conceptually different from each other.

Notwithstanding the above limitations, this research begins a much-needed discussion on the role of organizational capabilities on BDPA diffusion and its impact on OP.

7. Conclusions
The main purpose of this was to analyze the indirect impact of IT and HR capabilities on OP as their relationship is mediated through BDPA diffusion. We concluded that IT deployment and HR capabilities have a direct impact on BDPA diffusion, while these constructs have an indirect impact on OP. Our results also provide strong support for the full mediation effect of BDPA diffusion on the relation between the three IT deployment capabilities and OP while it partially mediates the relation between business–BDPA alignment, HR capabilities and OP. Based on these findings, we hope to contribute to the BDPA literature in identifying the role of organizational capabilities on BDPA diffusion and its impact on OP.
References


---

**MD**

---

**57,8**


Further reading


Corresponding author
Zongwei Luo can be contacted at: luozw@sustc.edu.cn

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Turning information quality into firm performance in the big data economy

Samuel Fosso Wamba
Department of Information Systems, Toulouse Business School, Toulouse, France

Shahriar Akter
Sydney Business School, University of Wollongong, Sydney, Australia, and
Laura Trinchera and Marc De Bourmont
Department of Information Systems, NEOMA Business School, Mont-Saint-Aignan, France

Abstract

Purpose – Big data analytics (BDA) increasingly provide value to firms for robust decision making and solving business problems. The purpose of this paper is to explore information quality dynamics in big data environment linking business value, user satisfaction and firm performance.

Design/methodology/approach – Drawing on the appraisal-emotional response-coping framework, the authors propose a theory on information quality dynamics that helps in achieving business value, user satisfaction and firm performance with big data strategy and implementation. Information quality from BDA is conceptualized as the antecedent to the emotional response (e.g. value and satisfaction) and coping (performance). Proposed information quality dynamics are tested using data collected from 302 business analysts across various organizations in France and the USA.

Findings – The findings suggest that information quality in BDA reflects four significant dimensions: completeness, currency, format and accuracy. The overall information quality has significant, positive impact on firm performance which is mediated by business value (e.g. transactional, strategic and transformational) and user satisfaction.

Research limitations/implications – On the one hand, this paper shows how to operationalize information quality, business value, satisfaction and firm performance in BDA using PLS-SEM. On the other hand, it proposes an REBUS-PLS algorithm to automatically detect three groups of users sharing the same behaviors when determining the information quality perceptions of BDA.

Practical implications – The study offers a set of determinants for information quality and business value in BDA projects, in order to support managers in their decision to enhance user satisfaction and firm performance.

Originality/value – The paper extends big data literature by offering an appraisal-emotional response-coping framework that is well fitted for information quality modeling on firm performance. The methodological novelty lies in embracing REBUS-PLS to handle unobserved heterogeneity in the sample.

Keywords User satisfaction, Firm performance

Paper type Research paper

Introduction

Big data has emerged as a new frontier for business in either establishing competitive advantages or exploiting untapped opportunities (Frisk and Bannister, 2017; Dubey et al., 2018; Prescott, 2014; Fosso Wamba et al., 2017; Akter et al., 2016; Hazen et al., 2014; El-Kassar and Singh, 2018). In every part of the world, industries and organizations collect more data than ever before, seeking smarter business strategies to harness this big data revolution. The extant literature identifies “big data” not only as “the next management revolution” (Mcafee and Brynjolfsson, 2012), but also as “the new raw material for business” (Economist, 2010), or “the new science that holds the answers” (Gelsinger, 2012). As it clearly appears in both the academic and practitioner literature, the increased attention to big data, and thus to big data analytics (BDA), is eloquent proof...
that the benefits of BDA are well acknowledged in any environment: better understanding of business, markets and consumers; higher productivity linked with profitability; and improved performance measurement mechanisms (Lavalle et al., 2011; Swafford et al., 2008; Mcafee and Brynjolfsson 2012; Elisabeth and Frank, 2017; Michael, 2014), amongst others. And all of these are constantly reflected in Google, Amazon, Harrah’s, Capital One, and Netflix’s business models. Companies aiming to leapfrog competition are increasingly interested in BDA to transform their business models, notably by customizing consumers’ desiderata, including when and how many they want, and what incentives will make them want more in their lifetime (Langenberg et al., 2012). However, despite the widespread buzz around BDA, leveraging BDA-driven information to generate business value continues to be a challenge for many organizations. This is why consulting firms such as Gartner, IBM and McKinsey & Co. have started providing services to help firms capitalize on this opportunity. The extant literature highlights that, “[a]s big data evolves, the architecture will develop into an information ecosystem: a network of internal and external services continuously sharing information, optimizing decisions, communicating results and generating new insights for businesses” (Sun and Jeyaraj, 2013). However, there are growing concerns and confusion regarding analytics-driven information quality (IQUL), business value (BVAL), user satisfaction (USAT) and firm performance (FPER) (Goes, 2014; Sun and Jeyaraj, 2013). Clearly, despite the paucity of research in this spectrum, a better understanding of IQUL dynamics is required in order to address the research gap. Because, “while generating quality information is the primary purpose of any IS [information system], few studies have explored the variables that affect Information Quality. This is a significant gap in the IS research. Quality information is a foundation of good decision making and positive outcomes, yet we know little about the variables that lead to improved Information Quality. More research is needed in order to understand better how to influence Information Quality” (Petter et al., 2013, p. 30).

In this study, we investigate ways to leverage IQUL in BDA so as to achieve enhanced firm performance, by proposing and testing a theory from the perspective of managers/users. This perspective is put in this context because firm performance ultimately depends on managers who are the most critical stakeholders, given their interest in knowing more about their businesses and therefore translating big data into better information and improved decisions (Mcafee and Brynjolfsson, 2012). The study also focuses on managers because they have the greatest curiosity about unlocking the power of big data for large-scale interventions and predictions (Davenport, 2012; Lavalle et al., 2011). Furthermore, the managers’ perspective is examined as they want to understand “how to fish out answers to important business questions from today’s tsunami of unstructured information” (Davenport and Patil, 2012, p. 73). Despite the importance of analytics-driven IQUL and its impact on USAT, BVAL and FPER, little research on manager-side BDA has focused on such dynamics. We aim to help fill this knowledge gap, and to this effect, we propose a conceptual model which is rooted in the traditional appraisal (IQUL)–emotional response (BVAL and USAT)–coping (FPER) framework (Lazarus, 1991; Michelman, 2017). To empirically test the proposed relationships, we collected data from 307 managers who rely on BDA for their day-to-day operations and strategic directions across various industries in the USA and France. The study’s findings suggest that analytics-driven IQUL has a positive impact on BVAL and USAT, which again influences FPER. Heterogeneity is likely to exist in the sample used in information systems (IS) studies (Becker et al., 2013). Therefore, we decided to investigate the presence of unobserved heterogeneity in our sample, thus coming out with three groups of business analytics users characterized by
different model parameters. More precisely, the study aims at answering the following research questions:

RQ1. How do IQUL perceptions of BDA determine critical business outcomes?

RQ2. Do existing groups of users share the same behaviors (in terms of strength of the effects) when determining the IQUL perceptions of BDA? And if yes, how different are they?

The answers to these research questions clearly contribute to the business–technology–analytics alignment of global organizations by framing the impact of IQUL on individual and business outcomes. This paper is structured as follows: the next section focuses on the conceptual model and the development of hypotheses, which is followed by the description of the adopted method and the research findings. The last section focuses on the study’s theoretical and practical contributions and provides guidelines for future research.

Research model
The proposed conceptual model on BDA illuminates IQUL as the core concept that enhances BVAL and USAT, which, in turn, influences FPER within an organization. The focus on analytics-driven IQUL to establish a linkage between BVAL, USAT and FPER is based on the fact that “big data still aims in large part to deliver the right information to the right person at the right time in the right form, but is now able to do so in a significantly more sophisticated form” (Agarwal and Dhar, 2014, p. 447). Using a coordination perspective, this study hypothesizes that IQUL enhances BVAL, which is required to increase USAT and the overall FPER. This investigation of a manager-side BDA strategy is set in analytics-driven organizations across various industries. The conceptual model draws on the IS and services marketing literature, thus enabling the interdisciplinary approach that is required to tackle the challenges and opportunities in BDA (Agarwal and Dhar, 2014; Goes, 2014). Figure 1 shows the research model while Table I defines the constructs in the model.

![Research model](image-url)
Defining big data analytics

Big data refers to huge quantities of data in the form of clickstreams, voices and videos, for transactions and other types of operations (Sun and Jeyaraj, 2013). In an attempt to define big data, Schroeck et al. (2012) identified its various dimensions, which span greater scope of information, real-time information, new kinds of data and analysis and non-traditional forms of media data, new technology-driven data, large volumes of data such as social media data, and the latest buzzwords. In their defining big data, IBM (2012), Johnson (2012), and Davenport (2013) focus more on aspects such as the variety of data sources, while other authors, such as Rouse (2011), Fisher et al. (2012), Havens et al. (2012), and Jacobs (2009), emphasize the importance of storing and analyzing "big data." IDC (2013) defines "big data" while focusing on its three main characteristics: the data itself, the analytics of the data, and the presentation of analytics results that allow business value creation in terms of new products or services. In this study, we define BDA as a holistic process that involves the collection, analysis, use and interpretation of data for various functional divisions, with a view to gaining actionable insights, creating business value, and establishing competitive advantages (Fosso Wamba et al., 2015).

Information quality

Drawing on coordination theories (Crowston, 1997; Malone and Crowston, 1990; Setia et al., 2013), this study proposes that BDA uses various sources of data to provide the business information that are needed to identify and assess patterns based on diverse actors. This diversity of data was highlighted in big data literature as, “[i]n fact, companies that learn to take advantage of big data will use real-time information from sensors, radio frequency identification and other identifying devices to understand their business environments at a more granular level, to create new products and services, and to respond to changes in usage patterns as they occur” (Sun and Jeyaraj, 2013). In other words, BDA can enable the coordination of data from a variety of fields to improve information quality and organizational performance. This study contends that complex and interdependent BDA platforms produce coordinated information for the enhancement of BVAL, USAT and FPER. The extant research assessing the organizational impacts of BDA highlights the importance of IQUL in these environments (Schläfke et al., 2013; Langenberg et al., 2012). The application of BDA-driven quality information, rather than gut instinct, in decision making has become a core focus of research after evidence of the success of FPER in many

<table>
<thead>
<tr>
<th>Table I. Constructs and definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information quality</td>
</tr>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Wixom and Todd (2005)</td>
</tr>
<tr>
<td>Gregor et al. (2006)</td>
</tr>
<tr>
<td>Spreng et al. (1996)</td>
</tr>
<tr>
<td>Miah et al. (2017) and Alan et al.</td>
</tr>
</tbody>
</table>

Turning information quality
organizations (Lavalle et al., 2011; McAfée and Brynjolfsson, 2012). The extant literature identifies that IQUL influences various outcomes, such as satisfaction (Nelson et al., 2005; Barney, 2001), loyalty (Zhou et al., 2009), trust in the IT artifact (Vance et al., 2008) and user and knowledge-sharing behavior (Durcikova and Gray, 2009). We propose that IQUL is a critical component of a firm’s BDA success (Delone and Mclean, 1992; Wixom and Todd, 2005). The ultimate managerial challenge in the BDA environment lies in the finding of patterns in data and their translation into useful business information as mentioned in big data literature; “but to compete on that information, companies must present it in standard formats, integrate it, store it in a data warehouse, and make it easily accessible to anyone and everyone” (Langenberg et al., 2012).

Information quality: the antecedent for generating business value and managers’ satisfaction in a big data environment
Organizations with BDA capabilities aim to establish a robust foundation of quality information for decision making and business problem solving (Wixom et al., 2013). BDA with high information quality facilitates intra-organization operational coordination, thus enhancing the effectiveness of functional managers and generating different types of business value, as reflected in Table II. The research model of this study is based on the appraisal-emotional response-coping framework (Lazarus, 1991; Michelman, 2017), which suggests that more cognitively oriented information quality and value appraisal lead to emotive satisfaction, which, in turn, drives firm performance. This study argues that the assessment of analytics-driven information and relevant business value (appraisal) results in an affective or emotional response (i.e. satisfaction), which again leads toward a coping behavior (firm performance). This situation is identified by Bagozzi as an “outcome desire fulfilment” in which a manager in a big data environment assesses information quality and business value to increase satisfaction, which, in turn, influences perceived firm performance.

This study focuses on IQUL dynamics because “quality information” is the primary purpose of any application of BDA; however, few studies have conceptualized BDA in this context. A recent review of IS success studies states that “[i]nformation is the core reason for IS, and Information Quality is particularly important to classes of IS related to business intelligence, data-driven decision making, among others. More research is needed in order to better understand how to positively influence Information Quality” (Petter et al., 2013, p. 43). Therefore, the proposed model addresses this gap by modeling the effects of IQUL on BVAL, USAT and FPER in the BDA context.

Information quality and business value
Business value is at the heart of what managers pursue from a BDA perspective. The extant literature reports that the business value of analytics will be directly influenced by information quality in a big data environment (Wixom et al., 2013). The importance of the relationship between IQUL and BVAL was evidenced by Lavalle et al.’s (2011) study ranging over 30 industries across 100 countries. This relationship is also highlighted because, “[t]he goal of big data programs should be to provide enough value to justify their continuation while exploring new capabilities and insights” (Mithas et al., 2013, p. 18). Drawing on Gregor et al. (2006), this study defines business value as having several dimensions, namely, transactional, strategic and transformational, all of which benefit from BDA. “Transactional value” refers to the benefits added to firms as a result of IT use through its support of operation management, thus improving efficiency and cutting costs (Levich, 2015). As shown by Davenport (2012), an alignment between analytics-driven information quality and operational effectiveness results in the identification of profitable
<table>
<thead>
<tr>
<th>Study</th>
<th>Organizational functions</th>
<th>Description</th>
<th>Firm(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davenport (2006)</td>
<td>Customer selection, loyalty, and service, Pricing, Product or service quality, Promotion, Sales, consumer research, and marketing</td>
<td>Identify customers with the greatest profit potential, loyalty, and service. Increase likelihood that they will want the product or service offering, retain their loyalty, Identify the price that will maximize yield or profit, Detect quality problems early and minimize them, Track the performance of every region, and marketing to improve total business performance by analyzing interrelationships among functional areas.</td>
<td>Harrah’s, Capital One, Barclays, Progressive, Marriott, Honda, Intel, Dell (DDB matrix), Procter &amp; Gamble (P&amp;G)</td>
</tr>
<tr>
<td>Sun and Jeyaraj (2013)</td>
<td>Pricing</td>
<td>Customer intelligence group examines usage patterns and complaints data to accurately predict customer defections.</td>
<td>United Parcel Service (UPS), Macy's.com</td>
</tr>
<tr>
<td>Schroek et al. (2012)</td>
<td>Pricing</td>
<td>Optimize pricing of 73 million items in just over one hour.</td>
<td>Automercesdos Plaza’s</td>
</tr>
<tr>
<td>DallMule and Davenport (2017)</td>
<td>Customer defection</td>
<td>Scheduling price reductions to sell perishable products before they spoil.</td>
<td>Royal Bank of Canada</td>
</tr>
<tr>
<td>Kiron et al. (2011)</td>
<td>Service innovation</td>
<td>Deriving the most accurate pricing of products and services with precise calculation of customer profitability.</td>
<td>Netflix</td>
</tr>
<tr>
<td>LaValle et al. (2011)</td>
<td>Pricing</td>
<td>Analyze customer choice and customer feedback from over one billion reviews.</td>
<td>Netflix</td>
</tr>
<tr>
<td>Manyika et al. (2011)</td>
<td>Data-driven customer insights</td>
<td>Use personal profile and psychology-based analytics to help people connect and fall into a loving relationship.</td>
<td>Match.com</td>
</tr>
<tr>
<td></td>
<td>New product development</td>
<td>Each new PayPal initiative across finance, operations, and products is examined with quantified impact and leveraging analytics.</td>
<td>PayPal</td>
</tr>
<tr>
<td></td>
<td>Market share analysis</td>
<td>Collected 80–90% of possibly needed information about customers to generate analytics-driven customer insights.</td>
<td>Best Buy</td>
</tr>
<tr>
<td></td>
<td>Direct marketing through recommendation, relationship marketing</td>
<td>Uses big data to capture market share from its local competitors and develop customer insights.</td>
<td>Tesco</td>
</tr>
<tr>
<td></td>
<td>Customer behavior, customer segmentation, customer profitability</td>
<td>Recommendation engine to generate “you might also want” prompts to generate sales.</td>
<td>Amazon.com</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Developed behavioral segmentation and a multi-tier membership reward program by analyzing customer profile, real-time changes in customer behavior, and customer profitability.</td>
<td>Neiman Marcus</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Organizational functions</th>
<th>Description</th>
<th>Firm(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chandrasekaran et al. (2013)</td>
<td>Customer segmentation, customer loyalty</td>
<td>Systematically integrates analytics and consumer insights using data from its Clubcard loyalty program to better segment and target customer occasions. Simulate new products placed on shelves in order to test design effects internally and with consumers to enhance product acceptability after launching.</td>
<td>Harrah’s, Progressive Insurance, Capital One</td>
</tr>
<tr>
<td></td>
<td>New product acceptance rate</td>
<td></td>
<td>Tesco</td>
</tr>
<tr>
<td>Davenport and Patil (2012)</td>
<td>(a) Core search</td>
<td>Google uses data scientists to refine its core search and ad-serving algorithms.</td>
<td>Procter &amp; Gamble (P&amp;G)</td>
</tr>
<tr>
<td></td>
<td>(b) Advertisements</td>
<td></td>
<td>Google</td>
</tr>
<tr>
<td></td>
<td>Product, feature (e.g. “People you may know”)</td>
<td>To generate ideas for products, features, and value-adding services. By using “People you may know,” they generated millions of new page views which resulted in LinkedIn’s growth trajectory shifting significantly upward.</td>
<td>LinkedIn</td>
</tr>
<tr>
<td></td>
<td>and value-adding service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liebowitz (2013)</td>
<td>Product management</td>
<td>Macy’s analyze data at stock-keeping unit (SKU) level to make sure of the ready availability of product assortments.</td>
<td>Macy’s.com</td>
</tr>
</tbody>
</table>
customers for Harrah’s, Capital One, and Barclays, and in yield maximization for Progressive and Marriott. In a similar spirit, Wixom et al. (2013) indicate that GUESS INC., a fashion retailer, has been able to use less paper, save time, reduce the number of meetings, and increase cycle time and convenience by embracing BDA.

“Strategic value” takes place when firms change either their strategy (the ways in which they operate) or their products through the use of BDA, with a view to gaining competitive advantages together with offering better products and services to customers than their competitors. As reported by Manyika et al. (2011), Amazon.com has been hugely successful in generating strategic business value by implementing BDA for direct marketing, using recommendations such as “you might also want” prompts. These authors also report that Neiman Marcus establishes competitive advantages in customer segmentation and targeting by analyzing their customer profile and real-time changes in customer behavior. Similar strategies have been applied by Harrah’s, Progressive Insurance, and Capital One, to personalize product offers and increase customer loyalty in a systematic and effective manner. The extant literature focuses on the strategic benefits of BDA, because “[o]ne important benefit is that users develop a deeper understanding of the business [...] this understanding led to better purchasing and distribution decisions, and, ultimately, more sales of higher profitability items” (Wixom et al., 2013, p. 118).

Finally, “transformational value” refers to the benefits which flow into organizations in many forms, such as offering firms a simplification of their business process by restructuring internal organizational processes and activities or by performing tasks in an innovative way (Madden, 2015; Steenbrugge et al., 2014; Kirac et al., 2015; Lue et al., 2014). BDA-driven information quality ensures “transformational value” by establishing a management culture based on factual and real-time decisions, a single version of truth, more collaboration, and the discovery of business patterns (Wixom et al., 2013). Although analytics-driven information quality plays a critical role in generating business value, there is a paucity of empirical studies which confirm this relationship in a big data environment (Wixom et al., 2013; Lavalle et al., 2011; Goes, 2014). Therefore, the study hypothesizes that:

H1. Perceived IQUL has a significant positive impact on perceived BVAL in BDA.

Information quality, business value and satisfaction
The extant literature in marketing (Kane, 2017; Bowers et al., 2017) and IS (Nelson et al., 2005; Wixom and Todd, 2006; Delone, 2003) identifies information quality as both a cognitive and attitudinal construct. In a big data environment, scholars (Langenberg et al., 2012; McAfee and Brynjolfsson, 2012) have demonstrated that user satisfaction has a significant impact on BDA use; that is, a higher level of satisfaction creates greater user dependence on BDA. An evaluation of managers’ (or users’) satisfaction can help to track areas for improvement in order to strengthen BDA systems. Thus, we postulate that:

H2. Perceived IQUL has a significant positive impact on perceived USAT in BDA.

H3. Perceived BVAL has a significant positive impact on perceived USAT in BDA.

Satisfaction and firm performance
In BDA, information quality is widely acknowledged as being vital for increasing business and firm performance (Wixom et al., 2013). The extant literature provides evidence of a relationship between satisfaction and firm performance in terms of return on investment (Anderson et al., 1994, 1997; Zeithaml, 2000); operating margin (Bolton, 1998; Rust et al., 1994, 1995); and profitability (Fornell et al., 2006, 2009; Mithas et al., 2013; Kane et al., 2017; Ransbotham and Kiron, 2017). In the context of healthcare, Srinivasan and Arunasalam (2013)
show that the application of BDA in the form of predictive analytics and text mining can benefit firms by reducing cost (i.e. reduced amount of waste and fraud) and improving the quality of care (i.e. safety and efficacy of treatment). Wixom et al. (2013) have demonstrated that BDA can improve firm performance by improving productivity in terms of tangible (i.e. less paper reporting) and intangible (company reputation) benefits. Thus, a firm that creates superior user satisfaction should be able to maximize firm performance by facilitating pervasive use and speed via insights from BDA. Following this reasoning, we put forward the following hypothesis:

H4. Perceived USAT has a significant positive impact on perceived FPER in BDA.

Business value and firm performance
According to the extant literature on BDA, the relationship between business value and firm performance appears as one of the key issues for potential investigation (Wixom et al., 2013; Mithas et al., 2013; Sharma et al., 2014; Agarwal and Dhar, 2014). The early research on IT business value focused on impact on organizational performance, which includes cost reduction, increased profitability, higher productivity, and competitive advantages (Devaraj and Kohli, 2000; Hitt and Brynjolfsson, 1996; Mukhopadhyay et al., 1995; Kiron, 2017). This study adopts the “proxy view of IT” in defining the business value of BDA, with indication of the individual perceptions of its usefulness or value through firm performance in financial units (Orlikowski and Iacono, 2001; Burns, 2014):

H5. Perceived BVAL has a significant positive impact on perceived FPER in BDA.

Measurement development
In this study, the US survey measurement items was developed using an approach similar to the one used by Wixom and Todd (2005) and proposed by Moore and Benbasat (1991). More precisely, all constructs as well as their items were drawn from prior literature and were then adapted to fit the business analytics context (Table III). Afterward, eight experienced IS academics went through the survey to ensure the content validity. The next step was a pilot testing of the questionnaire with a total of 52 respondents recruited from various business analytics groups on LinkedIn, following the same process that was used for the subsequent main survey (Newbert, 2007). A seven-point Likert scale was used for all our items.

Once the US version of the survey in English was validated, a process similar to the one used by Setia et al. (2013) was followed to translate the English version of the survey into French. This consisted of a professional translator translating the survey into French and then back into English to ensure the reliability of the translation. A bilingual member of the research team went through the two versions of the survey to validate the translation. A pre-test of the final French questionnaire with nine respondents was then realized to confirm the construct validity. Subsequently, the combined 61 respondents were used to assess the robustness of our proposed model.

Survey administration
The main survey for this study was administrated by a leading market research firm, and sampling and data collection were then achieved in France and the USA. The data collection for the two samples was conducted from April 4, 2014-April 17, 2014. For the French sample, an invitation to participate in the study was sent on April 4, 2014 to a random sample of 500 members of the French panel of business analysts, business analytics and IT professionals. In all, 337 panel members agreed to participate in the study. A reminder was sent to participants on April 10, 2014, and the survey was closed on April 17, 2014. After a careful analysis of all
<table>
<thead>
<tr>
<th>2nd-order constructs</th>
<th>Type</th>
<th>1st-order constructs</th>
<th>Type</th>
<th>Item labels</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information quality (Wixom and Todd, 2005)</td>
<td>Molecular</td>
<td>Completeness</td>
<td>Reflective</td>
<td>INFQ1</td>
<td>The business analytics used provide a complete set of information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>INFQ2</td>
<td>___ provide comprehensive information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>INFQ3</td>
<td>___ provide all the information needed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Currency</td>
<td>Reflective</td>
<td>INFQ4</td>
<td>___ provide the most recent information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>INFQ5</td>
<td>___ produce the most current information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>INFQ6</td>
<td>___ always provide up-to-date information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Format</td>
<td>Reflective</td>
<td>INFQ7</td>
<td>The information provided by the analytics is well formatted.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>INFQ8</td>
<td>___ well laid out.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>INFQ9</td>
<td>___ clearly presented on the screen.</td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
<td>Reflective</td>
<td>INFQ10</td>
<td>The business analytics used produce correct information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>INFQ11</td>
<td>___ provide few errors in the information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>INFQ12</td>
<td>___ provide accurate information.</td>
</tr>
<tr>
<td>Business value (Gregor et al., 2006)</td>
<td>Molecular</td>
<td>Transactional</td>
<td>Reflective</td>
<td>BVTN1</td>
<td>Savings in supply chain management.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVTN2</td>
<td>Reducing operating costs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVTN3</td>
<td>Reducing communication costs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVTN4</td>
<td>Avoiding the need to increase the workforce.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVTN5</td>
<td>Increasing return on financial assets.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVTN6</td>
<td>Enhancing employee productivity.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Strategic</td>
<td>Reflective</td>
<td>BVST1</td>
<td>Creating competitive advantage.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVST2</td>
<td>Aligning analytics with business strategy.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVST3</td>
<td>Establishing useful links with other organizations.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVST4</td>
<td>Enabling quicker response to change.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVST5</td>
<td>Improving customer relations.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVST6</td>
<td>Providing better products or services to customers.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transformational</td>
<td>Reflective</td>
<td>BVTR1</td>
<td>An improved skill level for employees.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVTR2</td>
<td>Developing new business plans.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVTR3</td>
<td>Expanding organizational capabilities.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVTR4</td>
<td>Improving business models.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>BVTR5</td>
<td>Improving organizational structure/processes.</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>2nd-order constructs</th>
<th>Type</th>
<th>1st-order constructs</th>
<th>Type</th>
<th>Item labels</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>na</td>
<td>na</td>
<td>Satisfaction (Spreng et al., 1996)</td>
<td>Reflective</td>
<td>SABA1</td>
<td>Overall, I am</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>SABA2</td>
<td>___ satisfied with business analytics</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>SABA3</td>
<td>___ contented with business analytics</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>SABA4</td>
<td>___ pleased with business analytics</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>FPBA1</td>
<td>___ delighted with business analytics</td>
</tr>
<tr>
<td>na</td>
<td>na</td>
<td>Firm performance (Miah et al., 2017)</td>
<td>Reflective</td>
<td>FPBA2</td>
<td>Using analytics improved</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reflective</td>
<td>FPBA3</td>
<td>___ during the last 2 years relative to competitors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>___ Customer retention</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>___ Sales growth</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>___ Profitability</td>
</tr>
</tbody>
</table>
responses, 150 valid questionnaires were considered correctly filled out and appropriate for further analysis. Thus, for the French sample, we had a response rate of 44.51 percent.

A similar process was used to collect data in the USA. More precisely, an invitation to participate in the study was sent on April 7, 2014 to a random sample of 826 members of the US panel of business analysts, business analytics and IT professionals. A total of 668 panel members agreed to participate in the study. A reminder was sent to participants on April 12, 2014, and the web-based questionnaire was closed on April 17, 2014. After a careful analysis of all responses, 152 valid questionnaires were considered correctly filled out and appropriate for further analysis. Therefore, for this study, we had a response rate of 22.75 percent, thus giving a final sample of 302 useful responses.

Data analysis
The proposed theoretical model includes two second-order latent constructs: IQUL measured by four first-order latent constructs, and BVAL measured by three first-order latent constructs. Overall, the model includes 11 latent constructs. The complexity of the proposed model, along with the hypothesis that model parameters may be affected by unobserved heterogeneity, renders the use of the partial least squares (PLS) path modeling (Wang et al., 2016) more appropriate to estimate the theoretical model (Peng and Lai, 2012). We applied the PLS path modeling (Wang et al., 2016) to estimate the theoretical model. According to Becker et al. (2013), unobserved heterogeneity may arise in an IS sample. This is particularly true in BDA, where it is unrealistic that a unique model may fits all the units.

We used the REBUS-PLS algorithm (Esposito Vinzi et al., 2008) to investigate the presence of unobserved heterogeneity in our sample. Recently, Becker et al. (2013) presented a modification of the original REBUS-PLS algorithm, that is, the PLS-POS algorithm. Both of these methods allow unobserved heterogeneity to be accounted for in the whole model (i.e. the measurement as well as the structural part). In comparison to the REBUS-PLS algorithm, the PLS-POS algorithm applies to both formative and reflective indicators. However, the PLS-POS algorithm requires the number of unobserved groups to be defined in the first place. When no prior information can be used to predefine the number of groups to detect, the analysis has to be run several times with a different number of groups. The solution that best fits the data is retained. However, in REBUS-PLS, the algorithm automatically detects the number of unobserved groups. This is a key advance when there is no information about the existence (and the number) of groups. Since our model only involves a reflective measurement model and no prior information was available on the number of groups to be used, we decided to apply the REBUS-PLS algorithm. The REBUS-PLS algorithm provides, at the same time, group membership for each respondent and group-specific model parameters.

Results and discussion
The REBUS-PLS algorithm is available in XLSTAT-PLS, version 2013.6.04. According to Aloysius et al. (2016), all item loading values higher than 0.70 are considered adequate. Moreover, composite reliability (CR) values higher than 0.70 are considered acceptable. For average variance extracted (AVE), a value that is higher than 0.50 is considered to be an acceptable measure justifying the use of a construct (Sun and Zhang, 2008).

Execution of the REBUS-PLS algorithm and measurement validation
The REBUS-PLS algorithm automatically detected three groups with similar size (G1, G2 and G3). More precisely, 98 respondents were included in the first group, G1 (i.e. 34 percent of the sample), 108 in the second group, G2 (i.e. 36 percent of the sample), and the remaining 96 respondents (i.e. 32 percent of the sample) in the third group (G3).
In addition, the CR was verified for all the constructs in both the global model and the local models (see Table IV) (Aloysius et al., 2016). All items, with the exception of the one associated with BVTR1 in the local model estimated for G2, were strongly loaded on the corresponding construct. Since the standardized loading associated with BVTR1 was higher than 0.8 in the other two groups and in the global model, we decided to retain it in the analysis. The AVE indexes were higher than 0.60 for all the constructs in the global and local models, thereby exceeding the threshold of 0.5 defined by Fornell and Larcker (1981). Discriminant validity, verified at the global model level as the square root of each AVE value (see Table IV), exceeded the inter-construct correlations in all the models (see Tables V–VIII) (Fornell and Larcker, 1981; Hillol and Viswanath, 2017; Daniel et al., 2017). However, the correlation between IQUL and BVAL exceeded the square root of the AVE associated with BVAL in the local models estimated for the groups 1 and 2 (see Tables VI and VII). Multicollinearity among the constructs was tested. Variance inflation factors (VIF) indexes were reported along with the structural model results in Table IX. All the VIF values were smaller than 10, thus indicating that no serious multicollinearity affected the structural models whether at the global or the local levels (Roden et al., 2017; Rashid et al., 2017; Sharma et al., 2009). The only VIF value exceeding the threshold of 5 (Noor et al., 2015) was the one measuring the multicollinearity between IQUL and BVAL for the prediction of USAT in G1 (Table IX). This was consistent with the discriminant validity results, indicating that IQUL and BVAL were more highly correlated for respondents in G1 than for all the other respondents.

The estimated local models differed based on the relationships in the structural model and on some of the mean values of the second-order constructs. Two-tailed t-tests with a Bonferroni correction were run to compare item and construct means across groups. In Table X, we report the mean values of all items at the aggregate and group levels. The results of the two-tailed Bonferroni tests for pairwise comparisons are presented in Table XI. According to the results reported in Tables X and XI, respondents in G2 showed higher item mean values than respondents in G1 and G3. This was particularly true for all items related to strategic and transformational aspects of BVAL and for those related to FPER.

Respondents in G3 had lower values for all items with the exception of the one related to the currency, format and accuracy aspects of IQUL. In particular, they had significant lower values for all the items associated with FPER. The main construct means are reported in Table XII. The results of pairwise comparisons among the construct means are reported in Table XIII. The mean values of all the constructs except for IQUL are significantly different across groups. In particular, G2 was characterized by significant, higher mean values for FPER and BVAL, while respondents in G1 were characterized by a significant, higher mean value for USAT. In accordance with the item mean values, G3 was characterized by the lowest mean values for all constructs. This was particularly true for FPER: respondents in G3 showed a mean value superior to one point (on a seven-point scale), but smaller than the other two groups (Table XII).

Moreover, post hoc analyses were run to characterize the REBUS-PLS-detected groups according to manager demographic characteristics, years of experience and firm size. For a given demographic variable, we computed the percentage of respondents showing a specific category (relative frequency per category (percent) in Table XV). We tested the difference between the relative frequencies among the groups by applying $\chi^2$ tests for proportion. Manager proportions among the groups were not significantly different with respect to the country of origin of respondents and the size of the firm where they were employed.

However, G3 was characterized by a significantly (at a level of significance of 0.05) higher percentage of female respondents than all other groups. Moreover, no respondent in this group had a primary qualification. As for G1, its proportion of young respondents (younger than 33 years old) was not significantly high, resulting in a group with less
<table>
<thead>
<tr>
<th>2nd-order LVs</th>
<th>1st-order LVs</th>
<th>Items</th>
<th>Global</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>Global</th>
<th>Composite reliability and AVE</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQUL</td>
<td>Completeness</td>
<td>INFQ1</td>
<td>0.932</td>
<td>0.975</td>
<td>0.901</td>
<td>0.927</td>
<td>CR: 0.939</td>
<td>CR: 0.960</td>
<td>CR: 0.916</td>
<td>CR: 0.905</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INFQ2</td>
<td>0.942</td>
<td>0.966</td>
<td>0.928</td>
<td>0.933</td>
<td>AVE: 0.838</td>
<td>AVE: 0.925</td>
<td>AVE: 0.784</td>
<td>AVE: 0.827</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INFQ3</td>
<td>0.871</td>
<td>0.945</td>
<td>0.824</td>
<td>0.886</td>
<td>CR: 0.934</td>
<td>CR: 0.973</td>
<td>CR: 0.886</td>
<td>CR: 0.900</td>
<td></td>
</tr>
<tr>
<td>Currency</td>
<td></td>
<td>INFQ4</td>
<td>0.915</td>
<td>0.970</td>
<td>0.889</td>
<td>0.903</td>
<td>AVE: 0.825</td>
<td>AVE: 0.949</td>
<td>AVE: 0.723</td>
<td>AVE: 0.816</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INFQ5</td>
<td>0.870</td>
<td>0.977</td>
<td>0.744</td>
<td>0.873</td>
<td>CR: 0.913</td>
<td>CR: 0.956</td>
<td>CR: 0.902</td>
<td>CR: 0.894</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INFQ6</td>
<td>0.940</td>
<td>0.976</td>
<td>0.908</td>
<td>0.932</td>
<td>AVE: 0.838</td>
<td>AVE: 0.925</td>
<td>AVE: 0.784</td>
<td>AVE: 0.827</td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td></td>
<td>INFQ7</td>
<td>0.953</td>
<td>0.979</td>
<td>0.941</td>
<td>0.947</td>
<td>CR: 0.957</td>
<td>CR: 0.974</td>
<td>CR: 0.940</td>
<td>CR: 0.950</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INFQ8</td>
<td>0.966</td>
<td>0.979</td>
<td>0.957</td>
<td>0.966</td>
<td>AVE: 0.825</td>
<td>AVE: 0.949</td>
<td>AVE: 0.814</td>
<td>AVE: 0.863</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td>INFQ9</td>
<td>0.948</td>
<td>0.975</td>
<td>0.950</td>
<td>0.924</td>
<td>CR: 0.957</td>
<td>CR: 0.974</td>
<td>CR: 0.940</td>
<td>CR: 0.950</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INFQ10</td>
<td>0.947</td>
<td>0.980</td>
<td>0.933</td>
<td>0.935</td>
<td>AVE: 0.881</td>
<td>AVE: 0.950</td>
<td>AVE: 0.842</td>
<td>AVE: 0.863</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INFQ11</td>
<td>0.937</td>
<td>0.978</td>
<td>0.904</td>
<td>0.932</td>
<td>CR: 0.934</td>
<td>CR: 0.973</td>
<td>CR: 0.902</td>
<td>CR: 0.919</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>INFQ12</td>
<td>0.931</td>
<td>0.966</td>
<td>0.917</td>
<td>0.919</td>
<td>AVE: 0.881</td>
<td>AVE: 0.950</td>
<td>AVE: 0.842</td>
<td>AVE: 0.863</td>
<td></td>
</tr>
<tr>
<td>BVAL</td>
<td>Transactional</td>
<td>BVTN1</td>
<td>0.833</td>
<td>0.928</td>
<td>0.712</td>
<td>0.773</td>
<td>CR: 0.941</td>
<td>CR: 0.973</td>
<td>CR: 0.902</td>
<td>CR: 0.919</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVTN2</td>
<td>0.893</td>
<td>0.937</td>
<td>0.826</td>
<td>0.872</td>
<td>AVE: 0.727</td>
<td>AVE: 0.870</td>
<td>AVE: 0.605</td>
<td>AVE: 0.656</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVTN3</td>
<td>0.865</td>
<td>0.918</td>
<td>0.814</td>
<td>0.826</td>
<td>CR: 0.950</td>
<td>CR: 0.975</td>
<td>CR: 0.918</td>
<td>CR: 0.922</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVTN4</td>
<td>0.821</td>
<td>0.906</td>
<td>0.731</td>
<td>0.843</td>
<td>AVE: 0.760</td>
<td>AVE: 0.888</td>
<td>AVE: 0.632</td>
<td>AVE: 0.665</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVTN5</td>
<td>0.858</td>
<td>0.940</td>
<td>0.809</td>
<td>0.771</td>
<td>CR: 0.950</td>
<td>CR: 0.975</td>
<td>CR: 0.918</td>
<td>CR: 0.922</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVTN6</td>
<td>0.843</td>
<td>0.936</td>
<td>0.768</td>
<td>0.767</td>
<td>AVE: 0.760</td>
<td>AVE: 0.888</td>
<td>AVE: 0.632</td>
<td>AVE: 0.665</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strategic</td>
<td>BVST1</td>
<td>0.863</td>
<td>0.960</td>
<td>0.807</td>
<td>0.766</td>
<td>CR: 0.950</td>
<td>CR: 0.975</td>
<td>CR: 0.918</td>
<td>CR: 0.922</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVST2</td>
<td>0.891</td>
<td>0.944</td>
<td>0.821</td>
<td>0.860</td>
<td>AVE: 0.760</td>
<td>AVE: 0.888</td>
<td>AVE: 0.632</td>
<td>AVE: 0.665</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVST3</td>
<td>0.817</td>
<td>0.942</td>
<td>0.724</td>
<td>0.723</td>
<td>CR: 0.950</td>
<td>CR: 0.975</td>
<td>CR: 0.918</td>
<td>CR: 0.922</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVST4</td>
<td>0.886</td>
<td>0.945</td>
<td>0.830</td>
<td>0.859</td>
<td>AVE: 0.760</td>
<td>AVE: 0.888</td>
<td>AVE: 0.632</td>
<td>AVE: 0.665</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVST5</td>
<td>0.894</td>
<td>0.950</td>
<td>0.866</td>
<td>0.823</td>
<td>CR: 0.950</td>
<td>CR: 0.975</td>
<td>CR: 0.918</td>
<td>CR: 0.922</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVST6</td>
<td>0.879</td>
<td>0.912</td>
<td>0.767</td>
<td>0.851</td>
<td>AVE: 0.760</td>
<td>AVE: 0.888</td>
<td>AVE: 0.632</td>
<td>AVE: 0.665</td>
<td></td>
</tr>
<tr>
<td>Transformation</td>
<td></td>
<td>BVTR1</td>
<td>0.839</td>
<td>0.966</td>
<td>0.593</td>
<td>0.801</td>
<td>CR: 0.949</td>
<td>CR: 0.979</td>
<td>CR: 0.889</td>
<td>CR: 0.930</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVTR2</td>
<td>0.884</td>
<td>0.933</td>
<td>0.789</td>
<td>0.826</td>
<td>AVE: 0.760</td>
<td>AVE: 0.924</td>
<td>AVE: 0.620</td>
<td>AVE: 0.728</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVTR3</td>
<td>0.905</td>
<td>0.951</td>
<td>0.838</td>
<td>0.874</td>
<td>CR: 0.950</td>
<td>CR: 0.979</td>
<td>CR: 0.918</td>
<td>CR: 0.922</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVTR4</td>
<td>0.918</td>
<td>0.969</td>
<td>0.863</td>
<td>0.900</td>
<td>AVE: 0.760</td>
<td>AVE: 0.924</td>
<td>AVE: 0.620</td>
<td>AVE: 0.728</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BVTR5</td>
<td>0.899</td>
<td>0.967</td>
<td>0.825</td>
<td>0.861</td>
<td>CR: 0.950</td>
<td>CR: 0.979</td>
<td>CR: 0.889</td>
<td>CR: 0.930</td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td></td>
<td>SABA1</td>
<td>0.861</td>
<td>0.957</td>
<td>0.888</td>
<td>0.884</td>
<td>CR: 0.950</td>
<td>CR: 0.939</td>
<td>CR: 0.889</td>
<td>CR: 0.930</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SABA2</td>
<td>0.913</td>
<td>0.966</td>
<td>0.888</td>
<td>0.884</td>
<td>AVE: 0.825</td>
<td>AVE: 0.921</td>
<td>AVE: 0.807</td>
<td>AVE: 0.732</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SABA3</td>
<td>0.918</td>
<td>0.965</td>
<td>0.907</td>
<td>0.852</td>
<td>AVE: 0.825</td>
<td>AVE: 0.921</td>
<td>AVE: 0.807</td>
<td>AVE: 0.732</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SABA4</td>
<td>0.890</td>
<td>0.931</td>
<td>0.881</td>
<td>0.793</td>
<td>CR: 0.950</td>
<td>CR: 0.979</td>
<td>CR: 0.889</td>
<td>CR: 0.930</td>
<td></td>
</tr>
<tr>
<td>Firm performance</td>
<td></td>
<td>FPBA1</td>
<td>0.896</td>
<td>0.957</td>
<td>0.826</td>
<td>0.808</td>
<td>CR: 0.945</td>
<td>CR: 0.973</td>
<td>CR: 0.914</td>
<td>CR: 0.881</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPBA2</td>
<td>0.899</td>
<td>0.953</td>
<td>0.872</td>
<td>0.780</td>
<td>AVE: 0.812</td>
<td>AVE: 0.924</td>
<td>AVE: 0.728</td>
<td>AVE: 0.640</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPBA3</td>
<td>0.910</td>
<td>0.963</td>
<td>0.853</td>
<td>0.824</td>
<td>CR: 0.950</td>
<td>CR: 0.979</td>
<td>CR: 0.889</td>
<td>CR: 0.930</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPBA4</td>
<td>0.901</td>
<td>0.972</td>
<td>0.859</td>
<td>0.809</td>
<td>AVE: 0.825</td>
<td>AVE: 0.921</td>
<td>AVE: 0.807</td>
<td>AVE: 0.732</td>
<td></td>
</tr>
</tbody>
</table>

Notes: CR, composite reliability; AVE, average variance extracted
experienced managers as compared to the other two groups. Regarding G2, it replicated a sample composition with all the demographic characteristics. However, it did not include managers lacking formal education.

The structural model

The results of the structural model testing are presented in Figure 2, and in Tables IX, XIV, XVI–XVIII. In Figure 2, we present the estimated structural path models at both the global model and group levels. The arrow thickness on the path depends on the associated significance at each path coefficient. As for the structural models, the three groups show different patterns of relationships among the second-order latent constructs: USAT and FPER (see Figure 2 and Tables IX and XIV). In general, the $R^2$ values of G1 are higher than those of other groups; it is also the group where the correlations among the latent constructs are higher (see Tables V–VIII). As our sample was of relatively small size (especially at local

| Table V. Correlation matrix among latent constructs in the global model |
|-----------------------------|---------------------|---------------------|---------------------|
| IQUL | 0.818 |
| BVAL | 0.779 0.815 |
| USAT | 0.744 0.757 0.908 |
| FPER | 0.682 0.809 0.666 0.901 |
| Note: The square roots of the average variance extracted (AVE) are shown on the diagonal (in italic) |

| Table VI. Correlation matrix among latent constructs in the local model for G1 |
|-----------------------------|---------------------|---------------------|---------------------|
| IQUL | 0.948 |
| BVAL | 0.947 0.924 |
| USAT | 0.929 0.889 0.960 |
| FPER | 0.922 0.931 0.860 0.961 |
| Note: The square roots of the average variance extracted (AVE) are shown on the diagonal (in italic) |

| Table VII. Correlation matrix among latent constructs in the local model for G2 |
|-----------------------------|---------------------|---------------------|---------------------|
| IQUL | 0.770 |
| BVAL | 0.734 0.713 |
| USAT | 0.721 0.861 0.898 |
| FPER | 0.676 0.813 0.796 0.853 |
| Note: The square roots of the average variance extracted (AVE) are shown on the diagonal (in italic) |

| Table VIII. Correlation matrix among latent constructs in the local model for G3 |
|-----------------------------|---------------------|---------------------|---------------------|
| IQUL | 0.765 |
| BVAL | 0.717 0.736 |
| USAT | 0.582 0.740 0.851 |
| FPER | 0.494 0.566 0.581 0.806 |
| Note: The square roots of the average variance extracted (AVE) are shown on the diagonal (in italic) |
### Table IX. Structural model results

<table>
<thead>
<tr>
<th>Dependent constructs</th>
<th>Structural paths</th>
<th>Standardized path coefficients</th>
<th>$R^2$ value</th>
<th>Contribution to $R^2$ (%) and $R^2$ value</th>
<th>VIF value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BVAL</td>
<td>IQUL → BVAL</td>
<td>0.779*** 0.947*** 0.734*** 0.717***</td>
<td>0.61 0.90 0.55 0.51</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>IQUL → USAT</td>
<td>0.393*** 0.843*** 0.191*** 0.106ns</td>
<td>0.63 0.86 0.76 0.55</td>
<td>46.19% 90.64% 18.19% 11.21%</td>
<td>2.520 9.73 3.915 2.04</td>
</tr>
<tr>
<td></td>
<td>BVAL → USAT</td>
<td>0.451*** 0.009ns 0.721*** 0.663***</td>
<td>0.66 0.87 0.70 0.38</td>
<td>87.51% 84.94% 57.88% 44.58%</td>
<td>2.337 4.78 2.206 2.22</td>
</tr>
<tr>
<td>USAT</td>
<td>BVAL → FPERF</td>
<td>0.716*** 0.795*** 0.496*** 0.290*</td>
<td>0.66 0.87 0.70 0.38</td>
<td>87.51% 84.94% 57.88% 44.58%</td>
<td>2.337 4.78 2.206 2.22</td>
</tr>
<tr>
<td>FPER</td>
<td>USAT → FPER</td>
<td>0.124* 0.153** 0.396*** 0.361**</td>
<td>0.66 0.87 0.70 0.38</td>
<td>12.49% 15.06% 42.12% 55.42%</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** *p-value < 0.05; **p-value < 0.01; ***p-value < 0.001
model level), we opted for using the traditional inference (i.e. t-test and p-value) to validate the significance of the model’s structural coefficients (Table IX). We also computed bootstrapped confidence intervals using \( n = 200 \) resamples (Table XVI). The results obtained are consistent with the significant coefficients obtained after correction for common method bias (Table XVIII). Each of the inner relationships is discussed below.

**Impact on business value.** Table IX shows that IQUL has a significant positive effect on BVAL for the global model and for all the three detected local models (G1, G2 and G3), thus supporting \( H1 \) for global model, G1, G2 and G3. According to Cohen (1988), and considering the \( f^2 \) values reported in Table XVII, IQUL has a large effect on BVAL at both the global and local levels. In addition, the impact of IQUL on BVAL is significantly higher for respondents in G1, as compared to the global model and the other local models, G2 and G3 (see Table XIV).

**Impact on satisfaction.** In the proposed model, we assumed that USAT would be explained by IQUL and BVAL. At the global model level, both IQUL and BVAL have
significant and moderate positive effect on USAT (Tables IX and XVII), thus validating $H2$ and $H3$ at the global level (Table XIX). Similarly, for respondents in G2, IQUL and BVAL still show significant positive effects on USAT (Table IX), thus validating $H2$ and $H3$ for G2 (Table XIX). However, for respondents in G2, the main driver of USAT is BVAL, which
contributes for about 82 percent of the explained variability, while IQUL only accounts for 8 percent of the explained variability (Table IX). Moreover, the effect of BVAL can be considerate as large according to the $f^2$ value in Table XVII (Cohen, 1988). Differences occur when comparing models estimated for respondents in G1 and G3 (Table XIV). For respondents in G3, BVAL is the only significant driver of USAT and it alone explains 55 percent of the variability of USAT ($R^2 = 0.55$) (Table IX) and shows a large effect on USAT according to the $f^2$ value in Table XVII, thereby validating only H3 for G3 (Table XIX). On the other hand, for respondents in G1, the only significant driver of USAT is IQUL: alone, it accounts for 86 percent of the variability of USAT ($R^2 = 0.86$) (Table IX) and shows a large effect on USAT (Table XVII), thus validating H2 for G3 (Table XIX). The non-significance of the coefficient linking BVAL to USAT in G1 may be due to the high correlation between the two independent variables; therefore, caution must be applied in interpreting this result. However, the VIF value associated with this structural relationship is smaller than 10 (Table IX), indicating that no serious multicollinearity affects the structural model for G1 (Roden et al., 2017).

Impact on firm performance. In the proposed model, we assumed that FPER would be directly dependent on BVAL and USAT. As shown in Table IX, the two exogenous variables have significant positive effects on FPER for all groups, and as a result, $H4$ and $H5$ are validated for the three groups, G1, G2 and G3, as well as for the global model (Table XIX). However, at the global model level and for respondents in G1, BVAL is the most important driver of FPER explaining 85 percent or more of the explained variability (Table IX). This is confirmed by observing the $f^2$ values in Table XVII: BVAL has a large effect on FPER, while USAT only shows a small effect on FPER.

This is not true for respondents in G2 and G3, for whom BVAL and USAT have similar impact on FPER. In particular, the effects of both BVAL and USAT are moderated for respondents in G2, while respondents in G3 seem to be more satisfaction-driven than those in G2 (Table XIV), even if BVAL and USAT have small effects on FPER.

Common method bias correction. Relations in the structural model may be inflated because of common method bias (Chin et al., 2012). To test for common method bias, we followed the approach proposed by Malhotra et al. (2006). We used the smallest observed correlation between the constructs (i.e. 0.328 equals to the correlation between FPER and IQUL) as the basis for our correction. The results show no significant differences among the constructs, indicating that the proposed model is not affected by common method bias.

### Table XIII.
Construct means comparison among REBUS Groups

<table>
<thead>
<tr>
<th>Mean comparison</th>
<th>IQUL</th>
<th>BVAL</th>
<th>USAT</th>
<th>FPER</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1 vs G2</td>
<td>0.104ns</td>
<td>0.442***</td>
<td>0.359*</td>
<td>0.386**</td>
</tr>
<tr>
<td>G1 vs G3</td>
<td>0.133ns</td>
<td>0.459***</td>
<td>0.416*</td>
<td>0.998***</td>
</tr>
<tr>
<td>G2 vs G3</td>
<td>0.216ns</td>
<td>0.902***</td>
<td>0.057ns</td>
<td>1.384***</td>
</tr>
</tbody>
</table>

Notes: Differences are expressed in absolute values. Significant differences are in italic. Bonferroni correction for multi-group comparison has been applied. *p-value < 0.05; **p-value < 0.01; ***p-value < 0.001

### Table XIV.
Structural model comparison among REBUS Groups

<table>
<thead>
<tr>
<th>Path coefficient comparison</th>
<th>BVAL</th>
<th>USAT</th>
<th>FPER</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1 vs G2</td>
<td>0.383**</td>
<td>0.669**</td>
<td>1.007**</td>
</tr>
<tr>
<td>G1 vs G3</td>
<td>0.272**</td>
<td>0.779**</td>
<td>0.562*</td>
</tr>
<tr>
<td>G2 vs G3</td>
<td>0.111ns</td>
<td>0.110ns</td>
<td>0.445*</td>
</tr>
</tbody>
</table>

Notes: Differences are expressed in absolute values. Significant differences are in italic. *p-value < 0.05; **p-value < 0.01; ***p-value < 0.001
completeness) as a proxy of common variance bias. We adjusted the correlations between the LVs for common variance bias and we used the adjusted correlations to estimate adjusted structural model parameters. The coefficients obtained after adjustment for common variance bias remained significantly different from zero (Table XVI), except for the coefficient linking SAT to FPER in the global model. This confirms the absence of common variance bias in our data and the robustness of our results.

Limitations
Prior to discussing the managerial and theoretical implications of this study, a number of limitations need to be recognized. First, the vast majority of items used for our constructs were measured using an anchored seven-point Likert scale ranging from “strongly disagree” (1) to “strongly agree” (7). This may introduce the so-called “acquiescence bias,” which is related to the “respondents’ tendency to respond to items positively without much regard for its true content” (Chin et al., 2008). Therefore, future studies may consider using the nine-point scale of fast form items with the two-anchor points ranging from −4 to +4 as

<table>
<thead>
<tr>
<th>Variable</th>
<th>Categories</th>
<th>Relative frequency per category (%)</th>
<th>Global n = 302</th>
<th>G1 n1 = 98</th>
<th>G2 n2 = 108</th>
<th>G3 n3 = 96</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>France</td>
<td>49.67</td>
<td>53.06</td>
<td>53.70</td>
<td>41.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>50.33</td>
<td>46.94</td>
<td>46.30</td>
<td>58.33</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>20.86</td>
<td>15.31</td>
<td>17.59</td>
<td>30.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>79.14</td>
<td>84.69</td>
<td>82.41</td>
<td>69.79</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>18–25</td>
<td>4.31</td>
<td>8.16</td>
<td>3.70</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26–33</td>
<td>17.22</td>
<td>23.47</td>
<td>12.04</td>
<td>16.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>34–41</td>
<td>28.48</td>
<td>23.47</td>
<td>28.70</td>
<td>33.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50 or more</td>
<td>25.83</td>
<td>20.41</td>
<td>28.70</td>
<td>28.13</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>No formal qualification</td>
<td>0.66</td>
<td>1.02</td>
<td>0.00</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary qualification</td>
<td>0.66</td>
<td>1.02</td>
<td>0.93</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Secondary qualification</td>
<td>5.30</td>
<td>4.08</td>
<td>6.48</td>
<td>5.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>College qualification</td>
<td>12.25</td>
<td>13.27</td>
<td>13.89</td>
<td>9.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undergraduate degree</td>
<td>30.13</td>
<td>25.51</td>
<td>31.48</td>
<td>33.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Postgraduate degree</td>
<td>50.99</td>
<td>55.10</td>
<td>47.22</td>
<td>51.04</td>
<td></td>
</tr>
<tr>
<td>Years of experience</td>
<td>Less than one year</td>
<td>5.96</td>
<td>8.16</td>
<td>3.70</td>
<td>6.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2–5</td>
<td>32.45</td>
<td>35.71</td>
<td>27.78</td>
<td>34.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6–10</td>
<td>19.21</td>
<td>21.43</td>
<td>19.44</td>
<td>16.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11–15</td>
<td>20.86</td>
<td>17.35</td>
<td>24.07</td>
<td>20.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16–20</td>
<td>9.93</td>
<td>11.22</td>
<td>11.11</td>
<td>7.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Over 20</td>
<td>11.59</td>
<td>6.12</td>
<td>13.89</td>
<td>14.58</td>
<td></td>
</tr>
<tr>
<td>Firm size</td>
<td>0–19</td>
<td>1.33</td>
<td>1.02</td>
<td>1.85</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20–99</td>
<td>3.97</td>
<td>3.06</td>
<td>3.70</td>
<td>5.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100–249</td>
<td>5.30</td>
<td>4.08</td>
<td>5.56</td>
<td>6.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250–499</td>
<td>6.29</td>
<td>5.10</td>
<td>6.48</td>
<td>7.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500–999</td>
<td>6.29</td>
<td>5.10</td>
<td>8.33</td>
<td>5.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1,000–2,499</td>
<td>9.27</td>
<td>10.20</td>
<td>9.26</td>
<td>8.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2,500–4,999</td>
<td>9.60</td>
<td>9.18</td>
<td>12.04</td>
<td>7.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5,000–9,999</td>
<td>9.93</td>
<td>8.16</td>
<td>9.26</td>
<td>12.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10,000–24,999</td>
<td>12.58</td>
<td>12.25</td>
<td>11.11</td>
<td>14.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25,000–49,999</td>
<td>5.63</td>
<td>3.06</td>
<td>5.56</td>
<td>8.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50,000–99,999</td>
<td>11.92</td>
<td>16.33</td>
<td>11.11</td>
<td>8.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100,000 or more</td>
<td>17.88</td>
<td>22.45</td>
<td>15.74</td>
<td>15.63</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Value displayed as percentage of total responses. Percentages that are significantly different from the others at level $\alpha = 0.05$ are in italic.
Bootstrap confidence interval obtained with $S = 200$ bootstrap samples

Global G1 G2 G3
Dependent constructs Structural paths Lower Upper Lower Upper Lower Upper Lower Upper
Business value IQ $\rightarrow$ BV 0.645 0.851 0.855 0.998 0.356 0.702 0.510 0.824
Satisfaction IQ $\rightarrow$ SAT 0.267 0.605 0.388 1.260 0.068 0.364 −0.120 0.242
BV $\rightarrow$ SAT 0.268 0.665 −0.337 0.547 0.847 1.319 0.445 0.893
Overall performance BV $\rightarrow$ Perf 0.669 0.982 0.596 0.970 0.284 0.902 −0.001 0.614
Sat $\rightarrow$ Perf 0.001 0.265 −0.068 0.322 0.054 0.566 0.032 0.700

**Table XVI.** The results of the bootstrap procedure

**Note:** Arrow thickness in the structural model is a function of the significance of the associated coefficient

<table>
<thead>
<tr>
<th>Dependent constructs</th>
<th>Structural paths</th>
<th>Global $R^2$</th>
<th>G1 $f^2$ values</th>
<th>G2 $f^2$ values</th>
<th>G3 $f^2$ values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business value</td>
<td>IQ $\rightarrow$ BV</td>
<td>1.552</td>
<td>8.729</td>
<td>1.206</td>
<td>1.042</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>IQ $\rightarrow$ SAT</td>
<td>0.164</td>
<td>0.537</td>
<td>0.059</td>
<td>0.011</td>
</tr>
<tr>
<td>BV $\rightarrow$ SAT</td>
<td>0.217</td>
<td>0.006</td>
<td>1.006</td>
<td>0.492</td>
<td></td>
</tr>
<tr>
<td>Overall performance</td>
<td>BV $\rightarrow$ Perf</td>
<td>0.652</td>
<td>1.017</td>
<td>0.194</td>
<td>0.066</td>
</tr>
<tr>
<td>Sat $\rightarrow$ Perf</td>
<td>0.019</td>
<td>0.038</td>
<td>0.125</td>
<td>0.093</td>
<td></td>
</tr>
</tbody>
</table>

**Table XVII.** The model's explanatory power and predictive validity of the model

**Notes:** Large effect sizes are in bold, small effect sizes are in italic
suggested by Chin et al. (2008). Second, the BDA-enabled improved firm performance cannot be fully assessed by a limited set of determinants. Therefore, further research might attempt to integrate more determinants including, for example, information quality with system quality (Wixom and Todd, 2005), or service quality with information quality (Barney, 2001). Third, this study measures the direct impact of a set of determinants of BDA on firm performance. Another area of future research may consist in looking at the first-order impact of BDA, which is the impact at the process level (Forbes, 2013; Mooney et al., 1996).

**Implications for practice**

From the managerial perspective, the following implications can be underscored. First, the study offers a set of determinants for business analytics that managers might use to assess the BDA potential within their organization. Second, the ability of the REBUS-PLS algorithm to automatically detect three distinctive groups of business analytics users may contribute to facilitating the design of IT features and interfaces that match each user group’s desires, thus fostering user acceptance and the use of IT systems. Third, the developed ability to identify distinctive user behavior groups within a sample may allow project stakeholders in charge of designing training programs and interventions to provide more targeted and personalized training to each group identified by the REBUS-PLS algorithm.

**Implications for research**

This study integrates constructs from Wixom and Todd (2005), Gregor et al. (2006), Spreng et al. (1996) and Tippins and Sohi to study the potential of BDA in enabling improved firm performance. However, unlike these earlier studies that investigated the relationship between the independent and dependent variables at the global level, the current study argues that the adoption behavior varies among adopters of any given IT artifact. Therefore, only assessing the importance of the relationship between independent and dependent variables at the global level does not capture these differences or the unobserved heterogeneity that exists in social data (Zhang and Wu, 2017). Consequently, this study uses

<table>
<thead>
<tr>
<th>Dependent constructs</th>
<th>Structural paths</th>
<th>Global</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business value</td>
<td>IQ→BV</td>
<td>0.671***</td>
<td>0.921***</td>
<td>0.605***</td>
<td>0.579***</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>IQ→SAT</td>
<td>0.348***</td>
<td>0.826***</td>
<td>0.165*</td>
<td>0.035ns</td>
</tr>
<tr>
<td>Overall performance</td>
<td>BV→Perf</td>
<td>0.668***</td>
<td>0.781***</td>
<td>0.459***</td>
<td>0.195****</td>
</tr>
<tr>
<td></td>
<td>Sat→Perf</td>
<td>0.076ns</td>
<td>0.130****</td>
<td>0.332**</td>
<td>0.258*</td>
</tr>
</tbody>
</table>

Notes: \( r_M = \) shared correlation resulting from CMB using the correlation between FPER and completeness as marker variable. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.1

**Table XVIII.** Path coefficients before and after correcting for CMB

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Global model</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>Supported</td>
<td>Supported</td>
</tr>
<tr>
<td>H2</td>
<td>Supported</td>
<td>Supported</td>
</tr>
<tr>
<td>H3</td>
<td>Supported</td>
<td>Not supported</td>
</tr>
<tr>
<td>H4</td>
<td>Supported</td>
<td>Supported</td>
</tr>
<tr>
<td>H5</td>
<td>Supported</td>
<td>Supported</td>
</tr>
</tbody>
</table>

**Table XIX.** Results of hypotheses testing
the REBUS-PLS algorithm, which is a response-based method, to capture this unobserved heterogeneity (Esposito Vinzi et al., 2008). In addition, this research work is a response to the call by Becker et al. (2013) for more studies to investigate unobserved heterogeneity more thoroughly. These authors actually found that over the last 20 years, the leading IS journals in the world had published very few articles having used a structural model in their research and having “examined unobserved heterogeneity.” In such articles, it was assumed that empirical data were homogeneous and represented a single population, and that this could lead to possible bias during the assessment of structural model parameters. Another implication triggered by our study is that, by applying the REBUS-PLS algorithm, it is possible to identify three groups of business analytics users (G1, G2, and G3), which are all characterized by different user’s behaviors (e.g. difference in values for structural model parameters). These results may facilitate the design of IT systems that fit each user’s behavior across each identified group, thus facilitating the adoption and use of the IT systems, as well as the extended use of the said IT systems.

In addition, this study provides some insights into the nature and role of IS quality, business value and satisfaction in creating improved firm performance through BDA, thus contributing to the emerging literature on BDA. Given the increased importance of business analytics in facilitating firm competitive advantage, future studies may build upon our proposed determinants to explore the potential of business analytics at the process, inter-organizational and societal levels (Chee et al., 2012; Singh and Gaur, 2017).

Conclusion
BDA have emerged as the new frontier of innovation and competition in the wide spectrum of the business landscape due to the challenges and opportunities created by the information revolution. BDA increasingly provide value to firms using the dynamics of information quality that transform data into practical insights for robust informed decision making and business problems solving. This is a holistic process which deals concurrently with data, sources, skills, and systems in order to create a competitive advantage. Leading e-commerce firms like Google, Amazon, and Facebook have already embraced BDA and experienced enormous growth. This study presents a useful starting point for understanding the IQUL dynamics in a big data environment, notably by modeling their impact on BVAL, USAT, and FPER. The study reflects that once BDA-driven IQUL is well understood and the identified challenges properly addressed, the BDA application will maximize business value, which facilitates pervasive usage and speedy delivery of insights across organizations.

References


Further reading


Corresponding author
Samuel Fosso Wamba can be contacted at: s.fosso-wamba@tbs-education.fr

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Identifying Industry 4.0 IoT enablers by integrated PCA-ISM-DEMATEL approach

Shubhangini Rajput and Surya Prakash Singh
Department of Management Studies,
Indian Institute of Technology Delhi, New Delhi, India

Abstract
Purpose – The purpose of this paper is to identify, analyze and model Internet of Things (IoT) enablers essential for the success of Industry 4.0.
Design/methodology/approach – IoT enablers for Industry 4.0 are identified from literature and inferable discussions with industry experts. Three different techniques namely, principal component analysis (PCA), interpretive structural modeling (ISM) and decision making trial and evaluation laboratory (DEMATEL) are applied to model IoT enablers. In addition to this, DEMATEL is also applied under two different situations representing the behavioral characteristic of experts involved. These are termed as optimistic (maximum) and pessimistic (minimum).
Findings – The integrated approach of PCA-ISM-DEMATEL shows that IoT ecosystem and IoT Big Data are the most influential or driving IoT enablers. These two enablers have been identified as the pillars for Industry 4.0. On the other side, IoT interchangeability, consumer IoT, IoT robustness and IoT interface and network capability have also been identified as the most dependent enablers for Industry 4.0.
Practical implications – The findings enable the industry practitioners to select the most appropriate driving enablers for an effective implementation of Industry 4.0.
Originality/value – The integrated approach-based hierarchical model and cause-effect relationship among IoT enablers are proposed which is a novel initiative for Industry 4.0. Moreover, two different variants of DEMATEL namely, pessimistic and optimistic are applied first time.
Keywords IoT, PCA, ISM, DEMATEL (pessimistic), DEMATEL (optimistic), Industry 4.0.

1. Introduction
In recent years, there have been great development and advances in Internet of Things (IoT) and its related areas such as big data, cloud computing and wireless technologies. These emerging trends would provide opportunities to enhance industrial upgradation and also visualizes the fourth industrial revolution, i.e., Industry 4.0. In literature, Industry 4.0 is also known as, “smart industry” or “smart manufacturing.” It came into existence in the year 2011 when German Government launched a project of digital manufacturing at the Hannover Messe (Wan et al., 2016). Its main goal includes automation, process improvement and productivity enhancement. This advancing new technology has paved its way of progression, which has not only reshaped but also transformed the process of manufacturing industry. Industry 4.0 is a combination of many emerging technologies originating from cyber-physical systems (CPS), IoT, industrial integration, cognitive computing, cloud computing and others (Xu et al., 2018). One of the basic components of Industry 4.0 is IoT, which leads to a paradigm shift for manufacturing companies. Many manufacturing companies are facing challenges in integrating basic components of Industry 4.0. IoT is a technological transformation of real-world data into virtual data and has the capability to share information, self-organize intelligently. IoT provides an environment where any object can be connected and communicated throughout the network. It is embedded in physical objects linked through networks and acts as an analytical tool to understand the complexity and respond promptly. IoT has been adapted by some of the organizations to aid in real-time information collection, which promotes the
organizational efficiency. IoT has also benefited the other areas related to logistics and supply chain (Tu, 2018). Industry 4.0 is an umbrella term for immense growth in information technology which has influenced productivity, accuracy, quality and efficiency of manufacturing companies. Reaching a higher level of production efficiency and productivity with increasing automation is the main goal of Industry 4.0. With the reinforcement of IoT, smart manufacturing is managed autonomously with maximum optimization of resources throughout the process.

The term IoT was first coined in the year 1999 by British technology pioneer Kevin Ashton (Madakam et al., 2015) to develop a system in which physical objects could be connected to the internet via sensors. He coined the term to determine the significance of radio-frequency identification (RFID) tags used in supply chains to the internet in order to track goods without any human intervention. IoT ecosystem consists of physical objects, which are connected and accessible over the internet. The “things” in IoT could be a person or any object with an embedded sensor with an allocated IP address that has the capacity to collect and exchange the information over the network without any human assistance (Madakam et al., 2015). The integrated technology in CPS not only enables its connection with the physical world for capturing humongous data but also ensures higher IoT security, and increased CPS efficiency and effectiveness.

IoT is a fast gaining momentum that can help industries increase process accuracy, precision, minimize cost and can perceive benefits from real-time information, which would help in perceiving informed decisions. IoT offers an opportunity to monitor actual performance and key performance indicators of an organization. This gives leverage to create their product-service systems and could unravel the potential of the systems innovativeness through IoT (Rymaszewska et al., 2017). This paper proposes pessimistic and optimistic approaches of decision making trial and evaluation laboratory (DEMATEL) that would administer a realistic assessment of the enablers. Therefore, the IoT enablers identified for Industry 4.0 would implement solutions to revamp the operation and business process of many manufacturing systems.

The paper is organized as follows. Section 2 presents the literature review of IoT. Section 3 provides description of the IoT factors identified from literature review. Section 4 describes the proposed principal component analysis, interpretive structural modeling (PCA-ISM)-DEMATEL integrated approach to model IoT enablers. Section 5 presents the modeling of IoT enablers. Section 6 illustrates the integrated analysis of PCA-ISM-DEMATEL followed by discussion and managerial implications in Sections 7 and 8, respectively.

2. Literature review

IoT transforms the physical objects into smart and intelligent objects. The concept of IoT is emerged from the concurrence of different fields and technological developments. It lies on the emergence of enabling functionalities that arises from technologies such as sensors and other communication technologies, i.e., actuators. In 2000, research was conducted on sensors, which have played a significant role in manufacturing, logistics and supply chain arena and has offered business benefit and interoperable feature under one system environment. Similarly, RFID (Violino, 2005; Ngai et al., 2008) is used to build up an IoT—a network that would allow companies to track goods through the supply chain and run many applications simultaneously. With the support of IoT, any industry can transform the business process and provides control under one infrastructure as well as offers opportunity to take informed and decentralized decision. IoT factors which supports Industry 4.0 and gives leverage to implement Industry 4.0 in any business organization has been identified from various authors and experts’ opinion.
Tan and Wang (2010) identified key features that are essential for IoT infrastructure namely, reliability, scalability, modularity, quality of service (QoS), integration and interoperability, interfacing and networking capabilities, security. Authors have stated that these essential features have emphasized on IoT system’s components for timely collection and delivering of information. However, they believed that security in IoT is a crucial issue, as it is an add-on feature and no technical solution for security and privacy of information is guaranteed.

Miorandi et al. (2012) have also identified some IoT requirements for its implementation such as heterogeneity, ubiquitous data exchange, monitoring, self-organization capabilities, security and privacy, interoperability. Authors believed that IoT system efficiency would maximize by increasing the output of the system in extreme conditions. But authors stated IoT is in need to maintain secure environment as transaction of large data takes place, and for this transaction protocol is built for different devices for efficient communication, hence security is an essentiality for IoT system.

Athreya and Tague (2013) laid emphasis on self-organization in the IoT to achieve network communication among devices, which allows monitoring the system’s functioning. In addition, they also identified key components for self-organization in IoT such as neighbor discovery, service recovery, integration and heterogeneity, which allows smooth functioning of the IoT system.

Borgia (2014) has determined IoT communication feature is defined in terms of reliability and secure connections in which different devices and components are networked and worked together for a given time period in a functioning environment. Author has also listed some key general features and requirements for IoT which are self-configure and routable, self-organized and adaptation, QoS, cost minimization, scalability, flexibility and security for information sharing and this information is supposed to follow the path between source and destination. For this, routing service quality is improved like network configuration, addressing failures, bandwidth delay so that cost could be minimized. Stankovic (2014) has mentioned security and big data as a problem and research challenge for detecting and diagnosing the attacks. Usman and Zhang (2014) have mentioned that integration and interoperability is a challenge for seamless communication in IoT. One of the important other features, i.e., compatibility is required for devices or components which should comply with the IoT system to operate on standards and maximizes the interoperability of the system.

Gupta (2015) has illustrated that as network nodes increases proportionately network dimension also increases. With the onset of IoT, more devices will be connected to the system, therefore monitoring would be a prime importance. Pozza et al. (2015) have identified components of IoT such as neighbor discovery and service recovery to identify node at a particular period but also to capture where or when node is expected to be available. Authors have also agreed that if system failure occurred then connection and services it was earlier providing would also lapse, so service recovery is an important parameter to be considered for IoT. Fersi (2015) has addressed extensibility as challenging issue that IoT system face while facilitating the integration and communication technologies. He also mentioned security and privacy as operational and infers that IoT should maintain its performance and should not loose data when system fails or left the network. He overviewed other basic features, which are mandatory for IoT system, i.e., transparency, real-time information, scalability, modularity and reliability. Wortmann and Flüchter (2015) stated that predictive maintenance and service recovery as an important factors. Authors signified that changes in the system are essential for functioning of the system, which support in recovering the services while reducing the risk of the system failure. Flexibility (Beigne et al., 2015) is a requirement for IoT devices for dynamic management of reprogrammable devices/components or any plug-in to increase platform capability.
Balte et al. (2015) quoted that due to various communication stacks and standards the traditional security services could not be applied to IoT system. For this, flexible security system needs to be invented which can manage security threats, malware attacks in a dynamic environment of IoT ecosystem.

Elkhodr et al. (2016) mentioned security as a challenge for authorization, authentication, bootstrapping and access control. Other main features such as integration and interoperability, monitoring was also considered for IoT system. Bagur (2016) has also mentioned integration and interoperability as a challenge for IoT system (www.blrlabs.com/casestudy/interoperability-and-iot).

Ahmed et al. (2017) stated that Big Data has its relevance in IoT as sensors generate large amount of data so it is required to manage and analyze data pattern. Besides this, connectivity, storage, QoS, real-time analytics and benchmark are the key requirements to improve IoT services through big data analytics. Lu (2017) has also addressed extensibility as challenging issue and it is examined as an important element for an IoT system to be extensible to allow new components and technologies to integrate IoT, which also induces systems viability.

El-Kassar and Singh (2018) mentioned that Big Data assists in decentralized decision-making processes and also offers comprehensive information. Therefore, Big Data has its requirement for organizational and environmental performance as well as to increase competitive advantage. The detailed concept and applications of Big Data can be referred from the work of Lamba and Singh (2017, 2018).

Self-adaptation (Weyns et al., 2018) is an approach to deal with uncertainties at run time. It provides a software system with feedback loop to collect data about environment and system to adapt itself to changes to offer quality objectives when required by the IoT system. An important aspect of IoT, i.e., consumer Internet of Things (CiOT) (www.i-scoop.eu/internet-of-things-guide/what-is-consumer-internet-of-things-ciot/) emphasizes on different consumer devices and applications, which drives to take decentralized decisions as well as increases context awareness among consumers. Shin (2017) stated that numerous developing industries are in need of the Big Data, IoT, Industry 4.0, artificial intelligence, and information and communication technology to develop strategic plans for innovation. This provides an industry to assess its capabilities to respond dynamic changes and delivers a new picture of the innovativeness. In fact, emerging industries are looking forward to develop the agile and sustainable IoT architecture to meet the current and future challenges of the business environment. Further details on innovation and sustainability toward designing Industry 4.0 can also be referred from Kaur et al. (2016), Singh and Gaur (2018), Singh (2018) and Dubey et al. (2018), respectively, which concludes that Industry 4.0 promotes maximizing accuracy and precision through better networking and connectivity using minimum human interventions. The IoT enabling factors for Industry 4.0 are discussed in the following section.

3. IoT factors description

Based on experts’ opinion and extensive literature review, 20 factors are identified. These 20 factors are later clustered into groups through PCA. Brief description of these 20 factors is provided here. However, the detailed description can be referred from the respective papers cited in the literature review section:

(1) Heterogeneity: IoT is distinguished by heterogeneity with respect to various devices, technologies and services which are integrated into the system and are expected to deliver their capabilities in terms of communication and computational aspects.

(2) Scalability: IoT system is identified by its scalability of systems, process or network to manage voluminous data or resources. Scalability is featured as if the system has
the capability to increase its total output under extreme load when any hardware or component is proportionally added.

(3) Ubiquitous data exchange: in IoT, significant role is played by wireless communication technologies which enable different smart devices or components to come into network for data exchange in different formats and in large amount.

(4) Monitoring: it assists in tracking the IoT enabled devices and data which are connected to the network.

(5) Self-organized and adaptation: this feature enables IoT system to retrieve data from different devices in required format, and for monitoring the ecosystem’s functioning and adjusting its behavior in response to the system itself.

(6) Self-configure and routable: it co-ordinates with other networked IoT devices for information routing between source and destination and contribute in network configuration and asynchronous transaction support.

(7) Compatibility and reliability: it keeps the system’s ability to work together with various different devices/components for specified period in a given functioning environment without implementing any changes to the system.

(8) Real-time information: it is important for timely collecting and delivering the information captured through RFID or sensors.

(9) Big data characteristics: an increase in the number of devices over the network with an exponential increase in data consumption gives leverage to big data and can assist in managing and analyzing the data.

(10) CIoT: it lies in the types of applications/devices and the technologies which drives consumers and their purpose. This assists consumers to take decentralized decisions as well as increases context enrichment and technology usage among consumers.

(11) Neighbor discovery and self-reaction: it detects nearer entities/components/services, thus, to bring them in the IoT system’s network for efficient and effective functioning and self-reaction to events and stimuli to which objects are subjected.

(12) Predictive maintenance and service recovery: it is essential to detect modifications in the condition of the system, i.e., updating, detecting errors and addressing failures to carry out maintenance services for increasing the viability of the system and would support service recovery as it depends on operations of the IoT system. If system failures occur, the network connections and services it was providing previously also dies.

(13) Transparency: it provides end to-end visibility of information and IoT system.

(14) Modularity: it emphasizes IoT system’s components for interchangeability.

(15) QoS: it enhances the quality for those services and applications which are characterized by real-time traffic (bandwidth, delay).

(16) Interfacing and networking capabilities: it enables RFID for interfacing with the physical world besides location and communication.

(17) Integration and interoperability: it integrates low power communication technologies to enhance system robustness as well as to develop semantics among devices connected in a network to provide data in standardized formats for efficient communication.
(18) Cost minimization: it optimizes IoT operational cost namely, maintenance, development, energy consumption or installation. With the support of sensors, equipment failure would minimize and allows the system to perform planned maintenance.

(19) Flexibility and extensibility: it ensures dynamic management of reprogrammable devices and components or any plug-in which maximizes the system's viability and platforms capability.

(20) Secure IoT ecosystem: it provides robustness to information privacy, authorization, integrity, authentication, trusted secure environment, compliance, security bootstrapping and confidentiality.

4. Proposed PCA-ISM-DEMATEL integrated framework

4.1 Principal component analysis (PCA)
This section provides brief overview of PCA, ISM and DEMATEL approaches, which are considered to propose an integrated framework of PCA-ISM-DEAMTEL. PCA is the foundation for multivariate data analysis. PCA technique is extensively analyzed in the literature and in-depth details can be referred from Wold et al. (1987). PCA analyses data table representing observations described by several dependent variables, which are inter-correlated. Its aim is to extract the information from the data sets and then express it as a set of new orthogonal variables called principal components. Similar to PCA, factor analysis (FA) is another technique developed by group of psychologists in the year 1930. However, PCA and FA are closely related to each other and are used for clustering factors.

4.2 Interpretive structural modeling
ISM was introduced by Warfield (1974). It is an interactive process in which a set of directly or indirectly related elements is structured into a comprehensive model. It aims to identify relationships among elements, which describe a problem. Its detailed methodology can be referred from the work of Warfield (1974) and some other works by Sushil (2012), Sohani and Sohani (2012), Khatwani et al. (2015) and Dubey and Singh (2015). Following are the basic steps for the development of ISM model:

- identification of variables;
- constructing structural self-interaction matrix (SSIM);
- deriving initial reachability matrix from SSIM;
- checking transitivity and constructing final reachability matrix;
- partitioning reachability matrix into different levels; and
- developing ISM model.

4.3 DEMATEL
DEMATEL is proposed by the science and human affairs program of the Battelle Memorial Institute of Geneva to solve interrelated and complex problems (Gabus and Fontela, 1973; Lin and Tzeng, 2009). This method builds the interrelationship between factors or criteria to build a network relationship map (Huang et al., 2007; Yang et al., 2008; Ou Yang, 2013; Ware et al., 2014). DEMATEL is a comprehensive method for designing and analyzing a structural model of causal–effect relationship (Wu and Lee, 2007). The DEMATEL method is applied using following steps namely, compute average initial direct-influence matrix, compute normalized initial direct-influence matrix, compute the total influence matrix, and
constructing the causal diagram. Its detailed methodology can also be referred from the work of Lamba and Singh (2018).

In addition to the basic DEMATEL, two variants of DEMATEL considering pessimistic and optimistic characteristic behavior of respondents are also applied in the paper.

4.4 Proposed integrated framework
The following section presents the proposed PCA-ISM-DEMATEL integrated framework. The proposed framework integrates the techniques of PCA, ISM and DEMATEL to identify key enablers of IoT, and also to show the relative importance of the enablers with their causal relationship in order to provide a reference for decision making (Figure 1).

5. Modeling IoT enablers for Industry 4.0
This section presents the modeling of IoT enablers for Industry 4.0 using three techniques (PCA, ISM and DEMATEL) discussed in the preceding section. The detailed demonstration of the proposed framework is described here.

5.1 IoT factors identification
In this step, factors influencing IoT for the success of Industry 4.0 are identified through literature review and expert opinion. Following are the 20 IoT factors:

1. heterogeneity;
2. scalability;

---

Figure 1.
Integrated PCA-ISM-DEMATEL IoT framework
(3) ubiquitous data exchange;
(4) monitoring;
(5) self-organized and adaptation;
(6) self-configure and routable;
(7) compatibility and reliability;
(8) real-time information;
(9) Big data characteristics;
(10) CIoT;
(11) neighbor discovery and self-reaction;
(12) predictive maintenance and service recovery;
(13) transparency;
(14) modularity;
(15) QoS;
(16) interfacing and networking capabilities;
(17) integration and interoperability;
(18) cost minimization;
(19) flexibility and extensibility; and
(20) secure IoT ecosystem.

5.2 Cluster formation using PCA
The steps involved in this methodology are as follows:

5.2.1 Data validation. For this study, the sample size is of 100 area experts and academicians. To conduct the reliable FA the sample size needs to big enough. For analysis of these data sets, excel and XLSTAT were used. The response record of the respondents is presented in Appendix 1.

5.2.2 Exploratory factor analysis. PCA using varimax rotation applied to the data sets where Kaiser-Meyer-Olkin (KMO) measures the sampling adequacy to provide the appropriateness of the data. Therefore the KMO value calculated for this data set is 0.747. Hence, the factors can be easily extracted.

5.2.3 Analysis of eigenvalues. Table I presents the eigenvalues resulting from the FA. It measures the amount of variation in the total sample accounted for by each factor. For F1 eigenvalue equals 3.515 and represents 17 percent of the total variability similarly for other factors the following result is obtained.

5.2.4 Correlations between variables and factors. Table II highlights the closely related variables and factors. The number of variables which bears the nearly same correlation value or closely related to each other are clustered under particular factor. The highlighted cells in the table are clustered and results in extraction of eight factors.

5.2.5 Correlation analysis. The following given map is called correlation circle (below on axes F1 and F2). It presents the correlation between a variable and principal component. It shows a projection of the initial variables in the factors space. When two factors are far from the center then they are relatively positively correlated and are close to each other, if they are orthogonal they are not correlated and are on the opposite side of the center then they are negatively correlated (Figure 2).
5.3 Hierarchical modeling using ISM

Various steps for the development of ISM model for Industry 4.0 are:

5.3.1 Identification of IoT enablers for Industry 4.0. Eight IoT enablers for the implementation of Industry 4.0 have been identified through literature review and discussions with area experts. These are: secure IoT system (F1), IoT interchangeability
(F2), IoT data mobility (F3), IoT Big Data (F4), IoT system capability (F5), CIoT (F6), IoT interface and network capability (F7) and IoT robustness (F8). The in-depth description of these enablers is discussed in Section 3.

5.3.2 Structural self-interaction matrix. Contextual relationship matrix among each pair of enablers is constructed. For analyzing the enablers in establishing SSIM, the following four notations have been used:

- V-enabler \(i\) influences enabler \(j\);
- enabler \(j\) influences enabler \(i\);
- X-enabler \(i\) and \(j\) influence each other; and
- O-enabler \(i\) and \(j\) are not related to each other.

For example: Table III.

5.3.3 Reachability matrix from the SSIM. From SSIM matrix, initial reachability matrix, i.e., binary matrix is derived by replacing V, A, X, O with 0 and 1 as per the following rules:

- if the \((i, j)\) entry in SSIM is V, then the \((i, j)\) entry in reachability matrix becomes 1 and the \((j, i)\) entry becomes 0;
- if the \((i, j)\) entry in SSIM is A, then the \((i, j)\) entry in reachability matrix becomes 0 and the \((j, i)\) entry becomes 1;
if the \((i, j)\) entry in SSIM is \(X\), then both the \((i, j)\) and \((j, i)\) entry in reachability matrix becomes 1; and

- if the \((i, j)\) entry in SSIM is \(O\), then the \((i, j)\) entry in reachability matrix becomes 0 and the \((j, i)\) entry becomes 0.

Applying the above rules to the SSIM matrix, an initial reachability matrix (Table IV) is constructed.

From this matrix, a final reachability matrix is derived, considering the transitivity rule, which implies that if a variable “A” is related to “B” and “B” is related to “C”, then “A” necessarily follows “C” (Table V).

5.3.4 **Partitions on the reachability matrix.** Once the final reachability matrix is constructed, it is partitioned into different levels through subsequent iterations. From this, reachability and antecedent set of each enabler is prepared. The reachability set of an enabler consists of an enabler itself and all other enablers which are influenced by it whereas the antecedent set of an enabler consists of itself and all the enablers which influence it.

<table>
<thead>
<tr>
<th>S. no.</th>
<th>Enablers</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IoT ecosystem</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>IoT interchangeability</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>IoT data mobility</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>IoT Big Data</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>IoT system capability</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>CloudIoT</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>IoT interface and network capability</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>IoT robustness</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table IV.
Initial reachability matrix

<table>
<thead>
<tr>
<th>S. no.</th>
<th>Enablers</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IoT ecosystem</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>IoT interchangeability</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>IoT data mobility</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>IoT Big Data</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>IoT system capability</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>CloudIoT</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>IoT interface and network capability</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>IoT robustness</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table V.
Final reachability matrix (including transitivity)

**Note:** \(^a\)values signify transitivity
Thereafter the intersection of these sets is achieved for all the enablers. The enablers for which the reachability and intersection sets are same, they occupy the highest level in the ISM hierarchy. The highest level enabler in the hierarchy is identified and is not considered for consecutive iterations. It would not be considered in influencing any other enabler above its own level. This same procedure is repeated to assign other enablers at different levels in the hierarchy. Further, these levels help in constructing the ISM model.

In Table VI, enabler IoT interface and network capability (F7) have same reachability and intersection sets. Therefore, it occupies the highest level in the hierarchy. The subsequent iterations are presented in the following tables (Tables VII and VIII).

5.3.5 Developing ISM model. The ISM model is constructed using final reachability matrix and the hierarchical level of enablers depicted in Table, respectively. The ISM model illustrates that IoT ecosystem (F1) and IoT Big Data (F4) are the most driving enablers Figure 3.

5.3.6 MICMAC analysis. This MICMAC analysis is also referred as Matriced’ Impacts croised-multiplication applique’ and classment (cross-matrix multiplication applied

<table>
<thead>
<tr>
<th>Enabler</th>
<th>Reachability set</th>
<th>Antecedent set</th>
<th>Intersection set</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 4</td>
<td>1, 4</td>
<td>VII</td>
</tr>
<tr>
<td>2</td>
<td>2, 6, 7</td>
<td>1, 2, 3, 4, 5</td>
<td>2, 6</td>
<td>IV</td>
</tr>
<tr>
<td>3</td>
<td>2, 3, 5, 6, 7, 8</td>
<td>1, 3, 4, 5</td>
<td>3, 5</td>
<td>V</td>
</tr>
<tr>
<td>4</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 4</td>
<td>1, 4</td>
<td>VII</td>
</tr>
<tr>
<td>5</td>
<td>2, 3, 5, 6, 7, 8</td>
<td>1, 3, 4, 5</td>
<td>3, 5</td>
<td>V</td>
</tr>
<tr>
<td>6</td>
<td>2, 6, 7</td>
<td>1, 2, 3, 4, 5, 6</td>
<td>2, 6</td>
<td>III</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>7</td>
<td>I</td>
</tr>
<tr>
<td>8</td>
<td>7, 8</td>
<td>1, 2, 3, 4, 5, 6, 8</td>
<td>8</td>
<td>II</td>
</tr>
</tbody>
</table>

Table VI. Iteration 1

<table>
<thead>
<tr>
<th>Enabler</th>
<th>Reachability set</th>
<th>Antecedent set</th>
<th>Intersection set</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 4</td>
<td>1, 4</td>
<td>VII</td>
</tr>
<tr>
<td>2</td>
<td>2, 6, 7</td>
<td>1, 2, 3, 4, 5</td>
<td>2, 6</td>
<td>IV</td>
</tr>
<tr>
<td>3</td>
<td>2, 3, 5, 6, 7, 8</td>
<td>1, 3, 4, 5</td>
<td>3, 5</td>
<td>V</td>
</tr>
<tr>
<td>4</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 4</td>
<td>1, 4</td>
<td>VII</td>
</tr>
<tr>
<td>5</td>
<td>2, 3, 5, 6, 7, 8</td>
<td>1, 3, 4, 5</td>
<td>3, 5</td>
<td>V</td>
</tr>
<tr>
<td>6</td>
<td>2, 6, 7</td>
<td>1, 2, 3, 4, 5, 6</td>
<td>2, 6</td>
<td>III</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>7</td>
<td>I</td>
</tr>
<tr>
<td>8</td>
<td>7, 8</td>
<td>1, 2, 3, 4, 5, 6, 8</td>
<td>8</td>
<td>II</td>
</tr>
</tbody>
</table>

Table VII. Iteration 2–7

<table>
<thead>
<tr>
<th>Enabler</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>IoT interface and network capability (F7)</td>
<td>I</td>
</tr>
<tr>
<td>IoT robustness (F8)</td>
<td>II</td>
</tr>
<tr>
<td>CIoT (F6)</td>
<td>III</td>
</tr>
<tr>
<td>IoT interchangeability (F2)</td>
<td>IV</td>
</tr>
<tr>
<td>IoT data mobility (F3)</td>
<td>V</td>
</tr>
<tr>
<td>IoT system capability (F5)</td>
<td>VI</td>
</tr>
<tr>
<td>IoT ecosystem (F1)</td>
<td>VII</td>
</tr>
<tr>
<td>IoT Big Data (F4)</td>
<td>VII</td>
</tr>
</tbody>
</table>

Table VIII. Hierarchical level of enablers
to classification). Its objective is to classify the enablers in accordance with the driving power and dependence. This distinguishes the enablers into four categories: autonomous, dependent, linkage and independent:

1. Autonomous variables: these variables have weak driving power and weak dependence.
2. Dependent variables: the variables in this category have weak driving power but strong dependence.
3. Linkage: the variables in this category have strong driving power as well as strong dependence.
4. Independent variables: the variables in this category have strong driving power and strong dependence.

Based on the categorization it is illustrated that first cluster is autonomous enablers which are significantly aloof from the system. In this case there are no autonomous enablers. The second cluster is dependent enablers and the given figure represents that enabler IoT interchangeability (F2) and CIoT (F6) has driving power of 4 and dependence of 6 as well as enabler IoT interface and network capability (F7) and IoT robustness (F8) has driving power of 1 and 2, dependence of 8 and 7, respectively. Therefore, it is positioned at a place which conforms to their driving power and dependence. The third cluster is linkage and any effect on these enablers will probably affect other enablers as well. In this case, there are no linkage enablers. The fourth cluster are independent enablers which consists of enabler IoT ecosystem (F1) and enabler IoT Big Data (F4) both has a driving power of 8 and dependence of 2, besides these other enablers F3, F5, i.e., IoT data mobility and IoT system capability has a driving power of 6 and dependence of 4, accordingly, they are positioned at a place which corresponds their driving power and dependence as shown in Figure 4.
5.4 Causal–effect analysis using DEMATEL
This method classifies the variables into cause and effect groups and forms the structural model and analyzed further. Cause and effect relationship between variables is interpreted using three approaches namely, average, pessimistic (minimum) and optimistic (maximum) aggregation approach of DEMATEL. Following sections describes detailed process of cause and effect relationship using DEMATEL.

5.4.1 Causal–effect analysis using DEMATEL (Average). In this approach, experts’ opinions are aggregated using arithmetic mean. The calculations involved are shown in the following steps.

Experts’ opinion and calculate average matrix $A$. The first step in this process is to gather the experts’ opinion and develop the pair-wise contextual relationships between identified variables on a scale of 0–4 where, 0 represents ‘no influence’ and 4 represents the “very high influence”. An nxn non-negative matrix, $X_k = [x_{ij}^k]$ is formed for each expert. If there are $k$ experts participating in the evaluation process, with $1 \leq k \leq K$, and $n$ is the number of variables, then the average matrix $A$ can be constructed as:

$$A = [a_{ij}] = \frac{1}{K} \sum_{k=1}^{K} x_{ij}^k.$$  \hspace{1cm} (1)

Table IX demonstrates the initial direct average relation matrix derived from the above given formula.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Row sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>3.400</td>
<td>3.200</td>
<td>3.600</td>
<td>3.200</td>
<td>2.800</td>
<td>3.400</td>
<td>4.000</td>
<td>23.600</td>
</tr>
<tr>
<td>2</td>
<td>1.400</td>
<td>0.000</td>
<td>1.600</td>
<td>1.600</td>
<td>1.200</td>
<td>1.600</td>
<td>2.000</td>
<td>2.400</td>
<td>11.800</td>
</tr>
<tr>
<td>3</td>
<td>2.000</td>
<td>2.400</td>
<td>0.000</td>
<td>0.800</td>
<td>1.600</td>
<td>1.600</td>
<td>2.600</td>
<td>3.000</td>
<td>14.000</td>
</tr>
<tr>
<td>4</td>
<td>3.200</td>
<td>1.800</td>
<td>1.800</td>
<td>0.000</td>
<td>3.000</td>
<td>2.400</td>
<td>2.000</td>
<td>2.400</td>
<td>16.600</td>
</tr>
<tr>
<td>5</td>
<td>1.200</td>
<td>0.800</td>
<td>2.000</td>
<td>0.400</td>
<td>0.000</td>
<td>1.600</td>
<td>2.200</td>
<td>2.200</td>
<td>10.400</td>
</tr>
<tr>
<td>6</td>
<td>0.800</td>
<td>0.600</td>
<td>1.000</td>
<td>0.600</td>
<td>0.800</td>
<td>0.000</td>
<td>1.200</td>
<td>1.400</td>
<td>6.400</td>
</tr>
<tr>
<td>7</td>
<td>0.600</td>
<td>1.800</td>
<td>1.200</td>
<td>1.400</td>
<td>1.400</td>
<td>1.200</td>
<td>0.000</td>
<td>2.200</td>
<td>9.800</td>
</tr>
<tr>
<td>8</td>
<td>1.200</td>
<td>1.800</td>
<td>2.400</td>
<td>0.800</td>
<td>1.400</td>
<td>1.000</td>
<td>2.600</td>
<td>0.000</td>
<td>11.200</td>
</tr>
</tbody>
</table>

Table IX. Initial direct average relation matrix $A$

Figure 4. MICMAC diagram of IoT enablers for Industry 4.0 using ISM
Calculating the direct relation matrix $D$. The normalized direct relation matrix $D$ is obtained from the average matrix $A$. This is achieved by using the following equations (Table X):

$$D = A / s, \quad s > 0,$$

where:

$$s = \max_{i,j} \left[ \sum_{j=1}^{n} a_{ij}, \quad \max_{j} \sum_{i=1}^{n} a_{ij} \right].$$

Calculating total relation matrix $T$. Once the normalized direct relation matrix $D$ is obtained, the total relation matrix is calculated to realize the direct or indirect relationship between the variables. It is computed as:

$$T = D \left(1 - D^{-1}\right),$$

where, $I$ is an $n \times n$ identity matrix (Table XI).

Calculating the sum of rows and columns of Matrix $T$. In total relation matrix $T$, the sum of rows and the sum of columns are represented by vectors $D$ and $R$, respectively. Let $D_i$ be the sum of $i^{th}$ row in matrix $T$. The value of $D_i$ indicates the total given both direct and indirect effects, that variable $i$ has on other variables. Let $R_j$ be the sum of the $j^{th}$ column in matrix $T$. The value of $R_j$ shows the total received both direct and indirect effects, that all other variables have on variable $j$ (Table XII).

Constructing a cause and effect relationship diagram. From Table XII, the values of $(D+R)$ and $(D-R)$ are computed. When $i = j$, the sum $(D_i + R_j)$ shows the total effects given and received by factor $i$ which signifies the relative importance of the variable. If $(D_i + R_j)$ is positive, the influencer variable $i$ is a net cause, while if it is negative, factor $i$ is a net receiver.
The cause and effect relationship diagram is constructed by mapping the values of \((D+R, D-R)\) on x- and y-axes, respectively.

5.4.2 Causal–effect analysis using DEMATEL (pessimistic). In this approach, experts’ opinions are aggregated using minimum value. This approach follows the same procedure as shown previously and is also illustrated in Appendix 2. Table XIII presents the results obtained for the direct and indirect influences of the enablers. On the basis of the computed values of \((D+R)\) and \((D-R)\) causal diagram is constructed.

5.4.3 Causal–effect analysis using DEMATEL (optimistic). In this approach, experts’ opinions are aggregated using maximum value. Table XIV presents the results obtained for the direct and indirect influences of the enablers and preceding procedural results are shown in Appendix 3. On the basis of the computed values of \((D+R)\) and \((D-R)\) causal diagram is constructed.

5.5 Integrated analysis using PCA-ISM-DEMATEL

5.5.1 IoT clusters from PCA. PCA provides eight clusters based on correlation between associated variables and factors. From PCA analysis it is evident that variables-predictive

<table>
<thead>
<tr>
<th>S. no.</th>
<th>Enablers</th>
<th>D</th>
<th>R</th>
<th>D+R</th>
<th>D−R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IoT ecosystem (F1)</td>
<td>0.279</td>
<td>0.062</td>
<td>0.341</td>
<td>0.217</td>
</tr>
<tr>
<td>2</td>
<td>IoT interchangeability (F2)</td>
<td>0.074</td>
<td>0.097</td>
<td>0.172</td>
<td>−0.023</td>
</tr>
<tr>
<td>3</td>
<td>IoT data mobility (F3)</td>
<td>0.107</td>
<td>0.099</td>
<td>0.206</td>
<td>0.008</td>
</tr>
<tr>
<td>4</td>
<td>IoT Big data (F4)</td>
<td>0.146</td>
<td>0.059</td>
<td>0.205</td>
<td>0.087</td>
</tr>
<tr>
<td>5</td>
<td>IoT system capability (F5)</td>
<td>0.062</td>
<td>0.086</td>
<td>0.158</td>
<td>−0.034</td>
</tr>
<tr>
<td>6</td>
<td>CIoT (F6)</td>
<td>0.023</td>
<td>0.086</td>
<td>0.109</td>
<td>−0.062</td>
</tr>
<tr>
<td>7</td>
<td>IoT interface and network capability (F7)</td>
<td>0.051</td>
<td>0.142</td>
<td>0.194</td>
<td>−0.091</td>
</tr>
<tr>
<td>8</td>
<td>IoT robustness (F8)</td>
<td>0.069</td>
<td>0.171</td>
<td>0.240</td>
<td>−0.102</td>
</tr>
</tbody>
</table>

Table XII. The direct and indirect influences

<table>
<thead>
<tr>
<th>S. no.</th>
<th>Enablers</th>
<th>D</th>
<th>R</th>
<th>D+R</th>
<th>D−R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IoT ecosystem (F1)</td>
<td>0.193</td>
<td>0.005</td>
<td>0.198</td>
<td>0.188</td>
</tr>
<tr>
<td>2</td>
<td>IoT interchangeability (F2)</td>
<td>0.011</td>
<td>0.038</td>
<td>0.049</td>
<td>−0.026</td>
</tr>
<tr>
<td>3</td>
<td>IoT data mobility (F3)</td>
<td>0.035</td>
<td>0.046</td>
<td>0.080</td>
<td>−0.011</td>
</tr>
<tr>
<td>4</td>
<td>IoT Big data (F4)</td>
<td>0.042</td>
<td>0.025</td>
<td>0.066</td>
<td>0.017</td>
</tr>
<tr>
<td>5</td>
<td>IoT system capability (F5)</td>
<td>0.027</td>
<td>0.036</td>
<td>0.063</td>
<td>−0.010</td>
</tr>
<tr>
<td>6</td>
<td>CIoT (F6)</td>
<td>0.000</td>
<td>0.028</td>
<td>0.028</td>
<td>−0.028</td>
</tr>
<tr>
<td>7</td>
<td>IoT interface and network capability (F7)</td>
<td>0.007</td>
<td>0.071</td>
<td>0.079</td>
<td>−0.064</td>
</tr>
<tr>
<td>8</td>
<td>IoT robustness (F8)</td>
<td>0.018</td>
<td>0.084</td>
<td>0.102</td>
<td>−0.066</td>
</tr>
</tbody>
</table>

Table XIII. The direct and indirect influences

<table>
<thead>
<tr>
<th>S. no.</th>
<th>Enablers</th>
<th>D</th>
<th>R</th>
<th>D+R</th>
<th>D−R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IoT ecosystem (F1)</td>
<td>0.397</td>
<td>0.130</td>
<td>0.527</td>
<td>0.267</td>
</tr>
<tr>
<td>2</td>
<td>IoT interchangeability (F2)</td>
<td>0.169</td>
<td>0.198</td>
<td>0.367</td>
<td>−0.029</td>
</tr>
<tr>
<td>3</td>
<td>IoT data mobility (F3)</td>
<td>0.211</td>
<td>0.192</td>
<td>0.403</td>
<td>0.020</td>
</tr>
<tr>
<td>4</td>
<td>IoT Big data (F4)</td>
<td>0.250</td>
<td>0.119</td>
<td>0.369</td>
<td>0.130</td>
</tr>
<tr>
<td>5</td>
<td>IoT system capability (F5)</td>
<td>0.116</td>
<td>0.281</td>
<td>0.397</td>
<td>−0.165</td>
</tr>
<tr>
<td>6</td>
<td>CIoT (F6)</td>
<td>0.085</td>
<td>0.132</td>
<td>0.217</td>
<td>−0.047</td>
</tr>
<tr>
<td>7</td>
<td>IoT interface and network capability (F7)</td>
<td>0.184</td>
<td>0.254</td>
<td>0.438</td>
<td>−0.069</td>
</tr>
<tr>
<td>8</td>
<td>IoT robustness (F8)</td>
<td>0.213</td>
<td>0.320</td>
<td>0.533</td>
<td>−0.107</td>
</tr>
</tbody>
</table>

Table XIV. The direct and indirect influences
maintenance, transparency, QoS, integration and interoperability, flexibility and secure IoT ecosystem are clustered as factor 1 (F1) as they are nearly correlated to each other. Other variables such as heterogeneity of devices, self-organized and adaptation, modularity are clustered as factor 2 (F2). Factor 3 (F3) comprises of ubiquitous data exchange, monitoring, compatibility and reliability. variables scalability and Big Data, real-time information and Neighbor discovery and self-reaction are clustered as factor 4 (F4) and factor 5 (F5), respectively. Factor 6 (F6) and factor 7 (F7) has only one variable namely, CIoT, interfacing and networking Capabilities, respectively. At last, self-configure and routable, and cost minimization are clustered as Factor 8 (F8). These eight IoT enablers are termed as IoT enablers and these eight clusters comprises 20 IoT factors identified from literature review and experts opinion (refer to Section 2). The proposed PCA-ISM-DEMATEL integrated approach is applied on these eight factors.

From the correlation circle, it is determined that the variable with the most explained variance of 17.576 is called as F1 and plotted on horizontal axes, the second most-explanatory variable with 8.937 variance is called F2 and plotted on the vertical axis. Within this 2-dimension circle the initial 20 variables are projected in red onto this 2-dimensional factor space. The red lines of the variables are in the same direction pointing up and to the right and are close together namely, heterogeneity of devices, self-organized and adaptation, modularity, scalability, integration and interoperability, flexibility and extensibility, monitoring, interfacing and networking capabilities, secure IoT ecosystem and the remaining variables are pointing down and to the right Big Data, compatibility and reliability, ubiquitous data exchange, cost minimization, QoS, transparency, predictive maintenance and service recovery, and real-time information. Since almost all variables are pointing toward the right, they are correlating on this first principal component F1. The other two variables CIoT, and self-configure and routable are on the other hand and are negatively correlated to second principal component F2.

5.5.2 IoT enablers from ISM. The ISM model and MICMAC analysis classifies the enablers according to its hierarchical level, dependence and driving power. From MICMAC analysis, it is illustrated that there are no enablers in the autonomous cluster, hence no enabler is treated isolated in the system and should be taken into account for the consideration of Industry 4.0. The next cluster consists of dependent enablers, which have weak driving power and strong dependence. The enablers in this cluster are IoT interchangeability (F2), CIoT (F6), IoT interface and network capability (F7), and IoT robustness (F8). These enablers are crucial to the Industry 4.0 success and have strong dependence; hence, they required other enablers to drive them. The other is linkage cluster and has strong driving power and dependence. There are no enablers in this linkage cluster. The final cluster consists of independent enablers, which have high driving and low dependence. The enablers in this cluster are IoT ecosystem (F1), IoT Big Data (F4), IoT data mobility (F3), and IoT system capability (F5). These enablers play a significant role to drive entire IoT enablers in the system and absence of these enablers may not drive other IoT enablers for Industry 4.0.

5.5.3 IoT enablers from DEMATEL. IoT enablers from DEMATEL is identified from three different aggregation approaches of DEMATEL namely, DEMATEL (average), DEMATEL (pessimistic), and DEMATEL (optimistic) IoT enablers analysis is carried out using these three different approaches.

DEMATEL (Average). According to the DEMATEL (average) approach, the enablers are arranged based on their (D+R) values as shown in table. The enabler IoT ecosystem (F1) has the highest value i.e. 0.341 followed by F8 > F3 > F4 > F7 > F2 > F5 > F6. Considering the respective values of (D–R) the enablers namely, IoT ecosystem (F1), IoT data mobility (F3) and IoT Big Data (F4) are categorized as cause group and IoT interchangeability (F2),
IoT system capability (F5), CIoT (F6), IoT interface and network capability (F7) and IoT robustness (F8) are categorized as effect group.

The enablers in the cause group (have +ve D–R value) are of great importance and have significant repercussions on the whole system. Among all the enablers, IoT ecosystem (F1) has highest (D–R) value of 0.217 which implies that it has significant importance for IoT infrastructure and also possesses high (D+R) value which states that it can influence all other enablers. In addition, F1 is also highly influenced at the same time by other enablers due to its influence given index (D) of 0.279 being a higher value and influenced received index (R) of 0.062 being smaller value. Similarly, the enabler IoT Big Data has second highest (D–R) value of 0.087 but on other side in terms of (D+R) value it is comparatively higher due to its high influence on other enablers and also being more influenced by them. The next enabler in the cause group IoT data mobility (F3) with (D–R) score of 0.008 and higher (D+R) score of 0.206. Also, its influence given index (D) of 0.107 value is the third highest among all enablers. It suggests that this enabler greatly affects the IoT infrastructure as well as it can influence the Industry 4.0 implementation.

Enablers who fall in to an effect group (have –ve D–R value) tend to be easily influenced by other enablers. Among all the enablers in the effect group, IoT robustness (F8) has the least (D–R) value of −0.102. However, its (D+R) value of 0.240 and influence received index (R) 0.171 is highest among all the enablers. It signifies that this enabler is the most influenced by other enablers but is still important for the success of Industry 4.0 owing to its second highest (D+R) value. The enabler IoT interface and network capability (F7) is second in the priority list in the effect group with (D–R) score of −0.091. It has comparatively least (D+R) score of 0.194 and which can be attributed to its influence given (D) score of 0.051, which is second least among all enablers. The other enabler CIoT (F6) with (D–R) value of −0.062 and (D+R) value of 0.109 signifies that it places third in effect group. The enablers IoT system capability (F5) has (D–R) value of −0.034 and least (D+R) value of 0.158 with influence given (D) value of 0.062. IoT interchangeability (F2) has (D–R) value of −0.023 and significantly less (D+R) value of 0.172 with influence given (D) value of 0.074, which suggests that it has paved the way for the foundation of Industry 4.0 to deal in effective and efficient functioning of the system as well as its maximizes system capability to work with different devices in a functioning environment.

DEMATEL (pessimistic). The DEMATEL (pessimistic) approach, the enablers according to its (D+R) value are arranged as F1 > F8 > F3 > F7 > F4 > F5 > F2 > F6. Considering the respective values of (D–R) the enablers namely, IoT ecosystem (F1) and IoT Big Data (F4) are categorized as cause group and remaining IoT enablers are categorized as effect group. Among all the enablers, IoT ecosystem (F1) has the highest (D–R) value of 0.188 and high (D+R) value of 0.198, which implies that its significant importance can influence all other enablers as well as being more influenced at the same time by other enablers as depicted from its influence given index (D) of 0.193 being a higher value and influenced received index (R) of 0.005 is less among all enablers. The enabler IoT Big Data has second the highest (D–R) value of 0.017 but on other side in terms of (D+R) value it is comparatively higher due to its high influence on other enablers and also being more influenced by them.

Among all the enablers in the effect group, IoT robustness (F8) has the least (D–R) value of −0.066. However, its (D+R) value of 0.102 is the second highest and influence received index (R) 0.084 is highest among all enablers. It signifies that this enabler is the most influenced by other enablers but still has its significance for Industry 4.0. The enabler IoT interface and network capability (F7) is second in the priority list in the effect group with (D–R) score of −0.064. It has comparatively small (D+R) score of 0.079 and can be attributed to its influence given (D) score of 0.007 which is the second least among all enablers. The other enabler CIoT
 having (D−R) value of −0.028 and (D+R) value of 0.028 signifies that it is placed third in effect group with no influence given (D) index. IoT interchangeability (F2) has (D−R) value of −0.026 and significantly less (D+R) value of 0.049 with influence given (D) value of 0.111. Similarly, the enablers IoT data mobility (F3) has (D−R) value of −0.011 and the third highest (D+R) value of 0.080 with influence given (D) value of 0.035, and IoT system capability (F5) has (D−R) value of −0.011 and the least (D+R) value of 0.063 with influence given (D) value of 0.027.

DEMATEL (optimistic). According to optimistic approach, the enablers according to its (D+R) value is arranged as F8 > F1 > F7 > F3 > F5 > F4 > F2 > F6. Considering the respective values of (D−R) the enablers namely, IoT ecosystem (F1), IoT data mobility (F3) and IoT Big Data (F4) are categorized as cause group and remaining IoT enablers are categorized as effect group. Among all enablers, IoT ecosystem (F1) has the highest (D−R) value of 0.267 and the second highest (D+R) value of 0.527 with influence given index (D) of 0.397 and influenced received index (R) of value 0.130. This states that it has significant importance and can influence other enablers and can be influenced from other enablers. Similarly, the enabler IoT Big Data (F4) has the second highest (D−R) value of 0.130 and (D+R) value of 0.369. In the same way, the IoT data mobility (F3) has (D−R) value of 0.020 and higher (D+R) value of 0.403 with given influence index (D) of 0.211.

Among all the enablers in the effect group, IoT system capability (F5) has (D−R) value of −0.165 and comparatively higher (D+R) value of 0.397 with influence given (D) value of 0.116. This suggests that it is influenced by other enablers and can also be considered as an important enabler. The second enabler in the effect group is IoT robustness (F8) which has (D−R) score of −0.107 and higher value of 0.533 among all enablers and it is the highly influenced enabler. Similarly, IoT interface and network capability (F7) has the least (D−R) value of −0.069. However, its (D+R) value of 0.438 and influence received index (R) 0.254 is the third highest among all the enablers signifies that this enabler is influenced by other enablers. The enabler CIoT (F6) is most influenced with (D−R) value of −0.047, and (D+R) value of 0.217. In the same way, IoT Interchangeability (F2) has (D−R) value of −0.029 and significantly low (D+R) value of 0.367 with influence given (D) value of 0.169.

Based on the integrated analysis of ISM and all variants of DEMATEL approach the IoT driving enablers are identified and the common driving IoT enablers are selected finally. The discussion of the common IoT enablers is provided in the following section.

6. Discussion
This section presents the common driving enablers across the integrated techniques of PCA-ISM-DEMATEL. PCA is applied to form the components from 20 identified factors which are similar to each other. Hence, eight clusters are formed using this technique and used for further analysis. The proposed integrated PCA-ISM-DEMATEL framework demonstrates that IoT ecosystem (F1) and IoT Big Data (F4) are the most influential driving enablers. By applying proposed framework, it is evident that these identified enablers are the most favorable IoT drivers for Industry 4.0 in any organization. The positioning of these enablers can be further analyzed using MICMAC and causal relationship of DEMATEL (average, pessimistic and optimistic approaches) as depicted in Figures 5–7. Moreover, it is interpreted that without IoT ecosystem (F1), Industry 4.0 cannot be successfully implemented. Another important enabler, i.e., IoT Big Data (F4) is a necessity to offer data storage and data processing. It also manages unstructured data generating from the IoT devices. From the ISM analysis, it is illustrated that enablers IoT ecosystem (F1), IoT data mobility (F3), IoT Big Data (F4) and IoT system capability (F5) are obtained as driving enablers. However, the most common enablers obtained from the DEMATEL approach considering average, pessimistic and optimistic aggregation approaches are IoT ecosystem
(F1) and IoT Big Data (F4) which signifies that these are the most crucial enablers and are the main roots of IoT infrastructure. IoT data mobility (F3) enabler is identified by DEMATEL average and optimistic approaches. The other enablers obtained from ISM are IoT system capability (F5) and IoT data mobility (F3) which has its potential in detecting the nearby entities and devices so as to bring them in a network for efficient an effective communication and also induces the monitoring of connected devices and data. Due to its ability, CIoT (F6) is also easily driven as it is emphasized on different types of technologies and devices. With the help of these three common driving enablers, consumers can make informed decisions and more importantly influence technology usage of the IoT system.
The most common enablers, i.e., IoT ecosystem (F1) and IoT Big Data (F4) drive other IoT enablers in the system because they play a vital role in developing the protocols for the heterogeneous devices so that they can communicate with each other. Besides this, it also secures the information and maintains the authorization, integrity, compliance and confidentiality. It enhances the system’s service quality by updating, detecting and addressing failures such as network configuration, integration of communication technologies, bandwidth delays, etc. When large amount of data gets transferred from source to destination IoT Big Data (F4) becomes mandatory as it is required for data formatting, data analysis, data pattern identification, predictive analysis, etc., because each device or component requires real-time information and in their desired format.

Therefore, based on the analysis derived from the proposed integrated approach, these common driving enablers should be considered for the fruitful success of Industry 4.0 as it is transforming present industries into smart and automated industry. It provides significant importance to the connectedness of people, objects, machines and information and communication technologies for the effective and efficient management of complex business process. The following set presents the summary of the proposed integrated framework of PCA-ISM-DEMATEL approach:

Enabler set$_{ISM} = \{F1, F3, F4, F5\}, \quad (5)$

Enabler set$_{DEMATEL\text{-}aggregate} = \{F1, F3, F4\}, \quad (6)$

Enabler set$_{DEMATEL\text{-}pess} = \{F1, F4\}, \quad (7)$

Enabler set$_{DEMATEL\text{-}opt} = \{F1, F3, F4\}, \quad (8)$

Common enabler set = \{F1, F4\}. \quad (9)

Equations 5–8 provide the IoT enablers obtained from ISM, DEMATEL (average), DEMATEL (pessimistic) and DEMATEL (optimistic) approaches of the proposed framework. Finally, Equation 9 provides the most common IoT enablers common to all techniques.
7. Managerial implications and theoretical contributions

The present study provides the significance of IoT enablers for industry, society, environment and an individual user. IoT has its great effect on human and technological evolution at a small scale or large scale. With the emergence of IoT, consumers drastically interacts with their homes, cars and other appliances which drives their way to digital society and provides opportunity to drive down costs and increase competitiveness. The factors identified in the paper provide fundamental insights to the decision makers in considering the design of Industry 4.0. The paper provides two significant enablers namely, IoT ecosystem (F1) and IoT Big Data (F4) which can be considered by the industry practitioners to implement the Industry 4.0 concept in the industry. It is recommended that these two enablers should be given high priority in the design of Industry 4.0 as compared to others enablers. Though the paper identifies 20 IoT enablers for Industry 4.0, however, theoretically these two enablers found to be driving the rest of the enablers. This is also evident from the proposed integrated approach. Thus, these two enablers can act as building blocks for the successful implementation of Industry 4.0. Therefore, these enablers offers IoT adopters the ability to collect and analyze data in real-time which are required for Industry 4.0.

8. Conclusion

The primary focus of this paper is to establish a hierarchy of enablers and classifying them according to their driving power and dependence so that decision makers can focus on specific enablers rather than considering all enablers. The successful implementation of Industry 4.0 with IoT system requires focal point in many different fields such as automobile industry, supply chain networks, shipping, manufacturing, aerospace industry, etc. To this effect, eight enablers were considered for study based on literature review and experts’ opinion. The hierarchical leveling and causal–effect relationship among these enablers is established using ISM and DEMATEL. ISM is used to define the enablers as dependent or independent drivers while DEMATEL is used to categorize the enablers into cause and effect groups. It is ascertained that IoT ecosystem (F1) and IoT Big Data (F4) are emerged as the common driving enablers across all techniques applied. On the other side, IoT data mobility (F3), IoT system capability (F7), IoT interchangeability (F8), CIoT (F2), IoT robustness (F6) and IoT interface and network capability (F5) are the dependent enablers. For effective implementation of Industry 4.0, IoT ecosystem (F1) and IoT Big Data (F4) and other enablers should be considered for robust IoT infrastructure. The important fact observed from MICMAC analysis is that there are no enablers clustered as autonomous which states that all enablers are considered important for IoT study. However, the priority of enablers can be decided as per their driving power and dependence. The pessimistic and optimistic approaches of DEMATEL applied in this study push toward a more realistic assessment of the enablers. The industry practitioners working in the area of IoT would find useful understanding of the enablers from hierarchical model presented in this paper. For future research directions, IoT initiatives can be further linked to other domains such as logistics, procurement, sustainability, block chain and fog computing.

References


Further reading


### Appendix 1

**Respondents Heterogeneity of devices**

<table>
<thead>
<tr>
<th>Respondents</th>
<th>Heterogeneity of devices</th>
<th>IoT scalability</th>
<th>Ubiquitous data exchange</th>
<th>Monitoring</th>
<th>Self-organized and adaptation</th>
<th>Self-configure and routable</th>
<th>Compatibility and reliability</th>
<th>Real-time information</th>
<th>Big Data</th>
<th>Consumer IoT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>9</td>
<td>10</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>17</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>18</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>20</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>21</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>22</td>
<td>9</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>23</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>24</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>25</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>26</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>27</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>28</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>29</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>30</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>31</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

**Table AI.** Respondent's response for PCA

(continued)
<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MD</td>
<td>57,8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table AI.**
<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>68</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>69</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>70</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>71</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>72</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>73</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>74</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>75</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>76</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>77</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>78</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>79</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>80</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>81</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>82</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>83</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>84</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>85</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>86</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>87</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>88</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>89</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>90</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>91</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>92</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>93</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>94</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>95</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>96</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>97</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>98</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>99</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>
| 100 | 5 | 9 | 8 | 7 | 8 | 8 | 8 | 8 | 8 | 8 | (continued)
<table>
<thead>
<tr>
<th>Respondents</th>
<th>Neighbor discovery and self-reaction</th>
<th>Predictive maintenance and service recovery</th>
<th>Transparency</th>
<th>Modularity</th>
<th>Service quality</th>
<th>Interface and network capability</th>
<th>Integration and interoperability</th>
<th>Cost minimization</th>
<th>Flexibility and extensibility</th>
<th>IoT security</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>16</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>17</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>18</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>19</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>21</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>22</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>23</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>24</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>26</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>27</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>28</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>29</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>30</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>31</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

(continued)
Table AI.

|   | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
|   | 7  | 8  | 8  | 9  | 8  | 8  | 9  | 8  | 8  | 9  | 8  | 9  | 9  | 8  | 8  | 9  | 8  | 9  | 8  | 9  | 8  | 8  | 8  | 8  | 8  | 8  | 8  | 9  | 9  | 9  | 9  | 9  | 9  | 9  | 9  | 9  | 9  | 9  |
| 32|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |

Identifying Industry 4.0 IoT enablers
<table>
<thead>
<tr>
<th>Table A1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>68</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>69</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>70</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>71</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>72</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>73</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>74</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>75</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>76</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>77</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>78</td>
<td>9</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>79</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>80</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>81</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>82</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>83</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>84</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>85</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>86</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>87</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>88</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>89</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>90</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>91</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>92</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>93</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>94</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>95</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>96</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>97</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>98</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>99</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>100</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>
## Appendix 2. DEMATEL – Pessimistic approach calculation

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Row sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>2.000</td>
<td>3.000</td>
<td>4.000</td>
<td>21.000</td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>4.000</td>
</tr>
<tr>
<td>3</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>2.000</td>
<td>2.000</td>
<td>8.000</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>2.000</td>
<td>1.000</td>
<td>2.000</td>
<td>2.000</td>
<td>9.000</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>2.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>2.000</td>
<td>6.000</td>
</tr>
<tr>
<td>6</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>3.000</td>
</tr>
<tr>
<td>8</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>2.000</td>
<td>0.000</td>
<td>0.000</td>
<td>5.000</td>
</tr>
<tr>
<td>Column sum</td>
<td>2.000</td>
<td>7.000</td>
<td>8.000</td>
<td>4.000</td>
<td>6.000</td>
<td>7.000</td>
<td>11.000</td>
<td>11.000</td>
<td></td>
</tr>
</tbody>
</table>

### Table AII. Initial direct minimum relation matrix A

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.143</td>
<td>0.143</td>
<td>0.143</td>
<td>0.143</td>
<td>0.095</td>
<td>0.143</td>
<td>0.190</td>
</tr>
<tr>
<td>2</td>
<td>0.048</td>
<td>0.000</td>
<td>0.000</td>
<td>0.048</td>
<td>0.000</td>
<td>0.000</td>
<td>0.048</td>
<td>0.048</td>
</tr>
<tr>
<td>3</td>
<td>0.048</td>
<td>0.048</td>
<td>0.000</td>
<td>0.048</td>
<td>0.000</td>
<td>0.048</td>
<td>0.048</td>
<td>0.095</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.048</td>
<td>0.048</td>
<td>0.000</td>
<td>0.095</td>
<td>0.048</td>
<td>0.048</td>
<td>0.095</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.095</td>
<td>0.000</td>
<td>0.000</td>
<td>0.048</td>
<td>0.048</td>
<td>0.095</td>
</tr>
<tr>
<td>6</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>0.000</td>
<td>0.000</td>
<td>0.048</td>
<td>0.000</td>
<td>0.000</td>
<td>0.048</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>0.000</td>
<td>0.048</td>
<td>0.048</td>
<td>0.000</td>
<td>0.048</td>
<td>0.048</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### Table AIII. Normalized direct relation matrix D

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.026</td>
<td>0.027</td>
<td>0.022</td>
<td>0.024</td>
<td>0.014</td>
<td>0.031</td>
<td>0.048</td>
</tr>
<tr>
<td>2</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>3</td>
<td>0.002</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.003</td>
<td>0.012</td>
<td>0.011</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.003</td>
<td>0.003</td>
<td>0.000</td>
<td>0.010</td>
<td>0.003</td>
<td>0.012</td>
<td>0.011</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.010</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.003</td>
<td>0.010</td>
</tr>
<tr>
<td>6</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>0.000</td>
<td>0.002</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>0.000</td>
<td>0.003</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.010</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### Table AIV. Total relation matrix T

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### Identifying Industry 4.0 IoT enablers

---
Appendix 3. DEMATEL – Optimistic approach calculation

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Row sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>4.000</td>
<td>4.000</td>
<td>4.000</td>
<td>4.000</td>
<td>3.000</td>
<td>4.000</td>
<td>4.000</td>
<td>27.000</td>
</tr>
<tr>
<td>2</td>
<td>2.000</td>
<td>0.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>3.000</td>
<td>4.000</td>
<td>17.000</td>
</tr>
<tr>
<td>3</td>
<td>3.000</td>
<td>3.000</td>
<td>0.000</td>
<td>1.000</td>
<td>3.000</td>
<td>2.000</td>
<td>3.000</td>
<td>4.000</td>
<td>19.000</td>
</tr>
<tr>
<td>4</td>
<td>4.000</td>
<td>3.000</td>
<td>2.000</td>
<td>0.000</td>
<td>4.000</td>
<td>3.000</td>
<td>2.000</td>
<td>3.000</td>
<td>21.000</td>
</tr>
<tr>
<td>5</td>
<td>2.000</td>
<td>1.000</td>
<td>2.000</td>
<td>1.000</td>
<td>0.000</td>
<td>2.000</td>
<td>3.000</td>
<td>3.000</td>
<td>14.000</td>
</tr>
<tr>
<td>6</td>
<td>1.000</td>
<td>1.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>0.000</td>
<td>2.000</td>
<td>2.000</td>
<td>12.000</td>
</tr>
<tr>
<td>7</td>
<td>1.000</td>
<td>3.000</td>
<td>2.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>0.000</td>
<td>4.000</td>
<td>18.000</td>
</tr>
<tr>
<td>8</td>
<td>2.000</td>
<td>3.000</td>
<td>4.000</td>
<td>1.000</td>
<td>4.000</td>
<td>1.000</td>
<td>4.000</td>
<td>0.000</td>
<td>19.000</td>
</tr>
<tr>
<td>Column sum</td>
<td>15.000</td>
<td>18.000</td>
<td>18.000</td>
<td>14.000</td>
<td>22.000</td>
<td>15.000</td>
<td>21.000</td>
<td>24.000</td>
<td></td>
</tr>
</tbody>
</table>

**Table AV.**
Initial direct maximum relation matrix A

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.148</td>
<td>0.148</td>
<td>0.148</td>
<td>0.148</td>
<td>0.111</td>
<td>0.148</td>
<td>0.148</td>
</tr>
<tr>
<td>2</td>
<td>0.074</td>
<td>0.000</td>
<td>0.074</td>
<td>0.074</td>
<td>0.074</td>
<td>0.074</td>
<td>0.111</td>
<td>0.148</td>
</tr>
<tr>
<td>3</td>
<td>0.111</td>
<td>0.111</td>
<td>0.000</td>
<td>0.037</td>
<td>0.111</td>
<td>0.074</td>
<td>0.111</td>
<td>0.148</td>
</tr>
<tr>
<td>4</td>
<td>0.148</td>
<td>0.111</td>
<td>0.074</td>
<td>0.000</td>
<td>0.148</td>
<td>0.111</td>
<td>0.074</td>
<td>0.111</td>
</tr>
<tr>
<td>5</td>
<td>0.074</td>
<td>0.037</td>
<td>0.074</td>
<td>0.074</td>
<td>0.000</td>
<td>0.074</td>
<td>0.111</td>
<td>0.111</td>
</tr>
<tr>
<td>6</td>
<td>0.037</td>
<td>0.037</td>
<td>0.074</td>
<td>0.074</td>
<td>0.074</td>
<td>0.000</td>
<td>0.074</td>
<td>0.074</td>
</tr>
<tr>
<td>7</td>
<td>0.037</td>
<td>0.111</td>
<td>0.074</td>
<td>0.111</td>
<td>0.111</td>
<td>0.074</td>
<td>0.000</td>
<td>0.148</td>
</tr>
<tr>
<td>8</td>
<td>0.074</td>
<td>0.111</td>
<td>0.148</td>
<td>0.037</td>
<td>0.148</td>
<td>0.037</td>
<td>0.148</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Table AVI.**
Normalized direct relation matrix D

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.057</td>
<td>0.058</td>
<td>0.049</td>
<td>0.065</td>
<td>0.035</td>
<td>0.064</td>
<td>0.069</td>
</tr>
<tr>
<td>2</td>
<td>0.015</td>
<td>0.000</td>
<td>0.018</td>
<td>0.015</td>
<td>0.020</td>
<td>0.015</td>
<td>0.033</td>
<td>0.052</td>
</tr>
<tr>
<td>3</td>
<td>0.028</td>
<td>0.032</td>
<td>0.000</td>
<td>0.007</td>
<td>0.036</td>
<td>0.016</td>
<td>0.036</td>
<td>0.056</td>
</tr>
<tr>
<td>4</td>
<td>0.045</td>
<td>0.034</td>
<td>0.021</td>
<td>0.000</td>
<td>0.056</td>
<td>0.030</td>
<td>0.023</td>
<td>0.041</td>
</tr>
<tr>
<td>5</td>
<td>0.014</td>
<td>0.007</td>
<td>0.016</td>
<td>0.005</td>
<td>0.000</td>
<td>0.014</td>
<td>0.030</td>
<td>0.032</td>
</tr>
<tr>
<td>6</td>
<td>0.005</td>
<td>0.006</td>
<td>0.014</td>
<td>0.012</td>
<td>0.016</td>
<td>0.000</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td>7</td>
<td>0.007</td>
<td>0.030</td>
<td>0.018</td>
<td>0.026</td>
<td>0.035</td>
<td>0.016</td>
<td>0.000</td>
<td>0.053</td>
</tr>
<tr>
<td>8</td>
<td>0.016</td>
<td>0.032</td>
<td>0.047</td>
<td>0.067</td>
<td>0.052</td>
<td>0.007</td>
<td>0.052</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Table AVII.**
Total relation matrix T

**About the authors**
Shubhangini Rajput is a full time PhD Scholar at Department of Management Studies, IIT Delhi, India. She works in the area of Operations and her research interests are modeling operations management issues and sustainable operations. She works in the area of Industry 4.0 and Internet of Things.
Surya Prakash Singh is Associate Professor in the Department of Management Studies, IIT Delhi, India. He holds a PhD from IIT Kanpur and Post-Doctoral Fellow from NUS Singapore-MIT USA Alliance. His research interest lies in facility layout problems, heuristics and meta heuristics. His work has been published in leading international journals such as IJPR, LNCS, IJAMT, EJM, RBR, IJRTE and APMR. His biography appeared in Marquis, USA Who’s Who in Science and Engineering, December 2007, and Who’s Who in the World, November 2010. Recently, he has been awarded Young Outstanding Faculty Fellowship from IIT Delhi. Surya Prakash Singh is the corresponding author and can be contacted at: surya.singh@gmail.com

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Value creation through big data in emerging economies

The role of resource orchestration and entrepreneurial orientation

Jing Zeng and Zaheer Khan
Kent Business School, University of Kent, Canterbury, UK

Abstract

Purpose – The purpose of this paper is to examine how managers orchestrate, bundle and leverage resources from big data for value creation in emerging economies.

Design/methodology/approach – The authors grounded the theoretical framework in two perspectives: the resource management and entrepreneurial orientation (EO). The study utilizes an inductive, multiple-case research design to understand the process of creating value from big data.

Findings – The findings suggest that EO is vital through which companies based in emerging economies can create value through big data by bundling and orchestrating resources thus improving performance.

Originality/value – This is one of the first studies to have integrated resource orchestration theory and EO in the context of big data and explicate the utility of such theoretical integration in understanding the value creation strategies through big data in the context of emerging economies.

Keywords Value creation, Resource management, Entrepreneurial orientation, Emerging economies, Big data, Resource-based view

Paper type Research paper

1. Introduction

Recently, big data has attracted increasing interest due to its potential to enhance organizational performance and its vital role in knowledge management (George et al., 2014; McAfee and Brynjolfsson, 2012; Khan and Vorley, 2017; Rothberg and Erickson, 2017; Zeng and Glaister, 2018). Big data refers to extremely large amounts of both structured and unstructured data sets that may be analyzed computationally by means of techniques that are characterized by high volume, high velocity and high variety, which traditional data processing technology is unable to store, capture and analyze (Chen et al., 2012; McAfee and Brynjolfsson, 2012; Schonberger and Cukier, 2013; Laney, 2001).

Many organizations are exploring ways of deploying and harnessing such large-volume data to create and capture value (Davenport, 2013; Waller and Fawcett, 2013; Wamba et al., 2017). This information resource may eventually overshadow physical resources (e.g. capital and labor) as the main driving force that re-shapes the competitive advantage of a firm (Bharadwaj et al., 2013; Davenport, 2013; Chen et al., 2012; Davenport and Patil, 2012; Koutroumpis and Leiponen, 2013; George et al., 2014). The importance of big data has recently drawn great attention from management scholars, leading to a small but growing stream of literature on the factors influencing big data decision-making quality (e.g. Janssen et al., 2017; Frisk and Bannister, 2017; Perez-Martin et al., 2018); managerial capability to drive big data value creation (e.g. Zeng and Glaister, 2018); big data analytics from health care, marketing and supply chain perspective (e.g. Wang and Hajli, 2017; Erevelles et al., 2016; Dubey, Gunasekaran and Childe, 2018); big data and business strategy alignment and performance (Akter et al., 2016; Wamba et al., 2017; Dubey, Gunasekaran, Childe, Luo, Wamba, Roubaud and Foropon, 2018; Prescott, 2014).

However, relatively few studies have sought to understand how managers transform resources to create value (e.g. Adner and Helfat, 2003; Helfat and Winter, 2011; Helfat et al., 2007; Ndofor et al., 2011; Sirmon et al., 2007, 2011). The exceptions are those relatively few
studies that have sought to understand this process by emphasizing the resource management process (e.g. Sirmon et al., 2007) and accentuating the dynamic capability of the firm based on “asset orchestration” (e.g. Helfat et al., 2007). Sirmon et al. (2011) later extends our understanding of the resource-related management literature by providing a holistic and complementary view that addresses explicitly how managers’ actions/capabilities to transform resources lead to value creation. The existing research on big data conducted by management scholars has predominantly relied on the dynamic capabilities approach to explain the potential impact of big data (e.g. Dubey, Gunasekaran, Childe, Luo, Wamba, Roubaud and Foropon, 2018; Erevelles et al., 2016; Opresnik and Taisch, 2015; Zeng and Glaister, 2018; Braganza et al., 2017). However, additional theory development is required to add richness to our understanding of how managers orchestrate resources in dynamic environments – such as those associated with big data and analytics – to improve organizational performance (Helfat and Winter, 2011; Helfat et al., 2007; Sirmon et al., 2011; George et al., 2014).

Despite the contributions made by the existing studies, however, relatively limited research has examined how value can be generated from big data in different contexts; the case in point being emerging markets, which may not possess the key skills needed for bundling and orchestrating big data-related resources for value creation strategies. The existing research noted that relevant big data analytics skills are mostly confined to developed economies – such as the California bay area – and are not readily available (Tambe, 2014). Companies may find it difficult to mobilize resources during the process of big data-related resource orchestration. One potential way through which managers can mobilize and orchestrate big data resources for value creation can be through the development of an entrepreneurial orientation (EO) (e.g. Lumpkin and Dess, 1996, 2001; Covin and Lumpkin, 2011).

Scholars from the field of entrepreneurship have examined corporate performance by paying closer attention to a company’s underlying EO. An EO refers to a company’s strategic orientation that captures entrepreneurial behaviors in terms of risk taking, proactiveness and innovativeness (e.g. Lumpkin and Dess, 1996; Wiklund, 1999; Rosenbusch et al., 2013). An EO can be one of the important links in explaining how a company can bundle and orchestrate the resources for value creating strategies (e.g. Sirmon et al., 2011). Thereby, an EO can explain, in part, the managerial processes that enable some companies to stay ahead of the competition; this is because an EO facilitates company action based upon early signals from its internal and external environments (Lumpkin and Dess, 1996). So far, the existing research on big data has applied a dynamic capabilities-based approach to explain the impact of big data on performance; yet, the way organizations orchestrate and bundle resources can be complex in changing environments. Thus, an EO can play a vital role in turning big data-related resources into knowledge assets by bundling and orchestrating big data-related knowledge for value creation (Lumpkin and Dess, 1996; Covin and Lumpkin, 2011).

This is the context in which the present paper aims to understand how companies based in emerging economies generate value from big data and improve their performance. The study adds to the limited research that has examined the potential performance implications of big data in emerging economy contexts. Understanding how managers orchestrate, bundle and leverage resources from big data in emerging contexts has wider implications for research on big data and its implications for corporate performance in different settings (e.g. Akter et al., 2016; Jabbour et al., 2017; Wamba et al., 2017).

This study makes three key contributions to the existing literature on big data and its implications for company performance. First, we bring resource orchestration into the domain of big data and identify an EO as one of the key factors through which companies bundle and orchestrate the knowledge assets arising from big data. The existing research
highlighted that resources themselves may not create value for companies; companies need to have internal practices and methods suited to putting resources into innovative value creating strategies. EO is one such method by which companies can improve performance through big data.

Second, this study provides important insights in terms of empirically demonstrating the value of integrating resource orchestration and EO in explaining the performance implications of big data. This is an important contribution as most of the existing literature on resource orchestration is conceptual in nature (Sirmon et al., 2007, 2011). Additionally, we provide important insights from the important emerging economy of China and show how managers bundle and orchestrate resources and create value from big data. This is one of the first studies to integrate resource orchestration theory and EO in the context of big data and to empirically support the arguments it puts forward by examining case studies from China.

2. Conceptual development

2.1 Resource orchestration and value creation through big data

The resource-based view (RBV) suggests that resources, on their own, may not be sufficient to create value, but that companies need to put in place an appropriate organization in order to take advantage of hard to imitate value creating resources (cf. Barney, 1991, 2001). Sirmon et al. (2007) defined resource management as the comprehensive process of structuring, bundling and leveraging company resources with the purpose of creating value for customers and competitive advantages for the company itself. The resource management process involves the three sub-processes of structuring, bundling and leveraging.

Structuring contains those processes by which companies acquire, accumulate and divest themselves of those resources that are affected by the environmental context. “Acquiring” refers to purchasing resources from strategic markets. “Accumulating” denotes the internal development of resources. “Divesting” pertains to the assessment – crucial for a company – of its existing resources, ridding itself of less-valued ones to generate the slack and flexibility needed to acquire and accumulate others of higher value (Sirmon and Hitt, 2003; Sirmon et al., 2007; Uhlenbruck et al., 2003).

Bundling includes: “stabilizing,” by which companies make minor incremental improvements to existing capabilities; “enriching,” which entails extending and elaborating current capabilities; and “pioneering”, which involves creating new capabilities.

Leveraging refers to the processes used to exploit a company’s capabilities and take advantage of specific market opportunities. According to Sirmon et al. (2007), effective cross-market leveraging capabilities include mobilizing, coordinating and deploying. “Mobilizing” refers to the capabilities required to form requisite capability configurations. “Coordinating” involves integrating capability configurations. “Deploying” involves physically using capability configurations to support the chosen leveraging strategy formed by the coordinating sub-process. Sirmon et al. (2007) further noted that, while each process and its sub-processes are important in themselves, they need to be synchronized in order to optimize value creation.

The process of resource management is referred to as managerial capabilities (Kraaijenbrink et al., 2010). Scholars have examined the managerial actions that affect resource management which in turn affect firms’ performance (Ndofor et al., 2011; Morrow et al., 2007), and the relationships that exist among resource management processes (Holcomb et al., 2009; Kor and Leblebici, 2005; Sirmon et al., 2011). These empirical studies have produced some important results. For example, Holcomb et al. (2009) indicated that the effects of managing resources are contingent on the quality of the resources held and on the synchronization of the processes used to manage them.
In parallel with research on the development of resource management, Helfat et al. (2007) put forward a related logic that focused on the asset orchestration emerging from the dynamic capability literature. Dynamic capability is an extension of RBV by highlighting explicitly the role of managers when they “purposefully create, extend or modify [the company’s] resource base” to create value to achieve sustainable advantage (Amit and Schoemaker, 1993; Eisenhardt and Martin, 2000; Teece et al., 1997; Winter, 2003). The recent work by Helfat and colleagues (Adner and Helfat, 2003; and Helfat et al., 2007) elaborated the concept of dynamic capability by accentuating the manager’s capabilities and decisions in influencing company’s performance regardless of the environment in which it operates. The asset orchestration, proposed by Helfat et al. (2007), consists of two primary processes: search and selection, and configuration and deployment. The search/selection process refers to a manager’s capability to identify assets and design organizational structures for the company and create business models to capture opportunities. The configuration/deployment process entails the coordination of co-specialized assets in order to nurture innovation. Helfat et al. (2007) argued that achieving a “fit” between these dimensions is a primary function of effective management. Essentially, dynamic managerial capabilities are largely created by adding new knowledge to the company’s current knowledge stocks (Adner and Helfat, 2003).

Sirmon et al. (2011) further extended RBV and their previous work (Sirmon et al., 2007) on resource management by bridging two related frameworks: resource management and asset orchestration. Resource orchestration was subsequently proposed, explicitly articulating managerial actions aimed at orchestrating resources in ways that help companies create a competitive advantage (Sirmon et al., 2011). Further research focusing on resource orchestration, highlighted by Sirmon et al. (2011), could serve as a catalyst for related research on the flow of knowledge within an organization.

The research being conducted in the area of resource management and asset orchestration is promising and encouraging; however, our understanding of how managers orchestrate a company’s resources could be enhanced by applying it to a big data context. This is the case for two reasons. First, big data, as an information asset, is a non-rivalrous resource due to its self-generative nature (Glazer, 1991). Therefore, making a distinction between non-rivalrous and rivalrous resources and understanding the value creation process based on the former could provide a more robust explanation for resource orchestration networks. Second, to date, very few empirical studies have explicitly incorporated resource orchestration into the heart of their inquiries (e.g. Chirico et al., 2011; Chadwick et al., 2015; Ndofor et al., 2011; Wales et al., 2013). For the most part, these studies deductively tested the relationship between managerial actions in relation to the connection between resources and performance (e.g. Ndofor et al., 2011; Wales et al., 2013; Chadwick et al., 2015; El-Kassar and Singh, 2018), or in family-run company contexts (Chirico et al., 2011). Furthermore, the underlying methods and managerial actions for the orchestration of resources are neither well theorized nor empirically proven. Thus, there is a great opportunity to understand how big data-related resources are orchestrated and leveraged in different contexts (Akter et al., 2016; Dubey, Gunasekaran and Childe, 2018; Jabbour et al., 2017; Prescott, 2014; Wamba et al., 2017). Such examination will provide not only important insights, but also a theoretically rich understanding of resource orchestration in the context of big data.

2.2 Entrepreneurial orientation and the orchestration of big data-related knowledge assets

Resources, on their own, may not create value; companies need to have internal managerial processes, structures and strategies in place to take advantage of resources and capture value from difficult to imitate resources (Barney, 1991; Eisenhardt and Martin, 2000). Due to its EO three set of characteristics of innovativeness, proactiveness and risk taking
(e.g. Wiklund, 1999; Covin and Lumpkin, 2011), an EO can be one of the important internal company-specific processes that can bundle and orchestrate knowledge assets originating from big data for pursuing innovative opportunities for the development of competitive advantage (Covin and Lumpkin, 2011; Prescott, 2014; Wamba et al., 2017).

The existing studies have examined the direct relationship between individual sets of resources and company performance; however, there has been relatively limited research focus on understanding how managers can effectively utilize those resources for value creation (Helfat, 2000). Since EO is often associated with a company’s strategic actions in capturing specific entrepreneurial aspects of decision-making styles, methods and practices, it is perceived by entrepreneurship scholars as one of the key capabilities that can explain the differential performances of companies (Lumpkin and Dess, 1996; Covin and Lumpkin, 2011). Applied to the resource orchestration framework, an EO may shed light how management can utilize and coordinate resources – such as big data-oriented ones – to improve performance (Simsek et al., 2010; Wales et al., 2013).

Resource bundling and orchestration in a big data environment is vitally important for achieving sustainable performance (e.g. Prescott, 2014; Dubey, Gunasekaran, Childe, Luo, Wamba, Roubaud and Foropon, 2018; Wamba et al., 2017). Due to its characteristics in terms of managerial practices and methods, an EO may play an important role in the orchestration of the resources, as managers will prepare the company to generate value from big data; this is because an EO “provides the mobilizing vision to use firm resources. By directing the use of resources, EO not only provides an objective, but also helps identify the resources necessary to support the objective” (Chirico et al., 2011, p. 311), as it refers to the “strategy making practices, management philosophies, and firm-level behaviors that are entrepreneurial in nature” (Anderson et al., 2009, p. 220). Thus, drawing insights from two sets of frameworks, an EO offers a complementary and integrated understanding of managerial actions in creating value from big data. Our objective, hence, is to add richness to current theory by extending the logic and ideas of resource orchestration to a company’s harnessing of big data (e.g. Akter et al., 2016; Dubey, Gunasekaran and Childe, 2018; Jabbour et al., 2017). In the following section, we elucidate our context, data collection and analysis procedures.

### 3. Context and research methods

Due to the paucity of research on resource orchestration and EO in the context of big data, we adopted an inductive, multiple-case research design that allows a “replication” logic (Yin, 2003) and in which cases are treated as experiments that confirm or refute the inferences drawn from others (Yin, 2014; Eisenhardt, 1989). This process typically creates opportunities to triangulate the information collected, augment external validity and help guard against observer bias, and yields more robust, generalizable theory than single cases (Eisenhardt and Graebner, 2007; Ketokivi and Choi, 2014; Miles and Huberman, 1994; Pagell and Wu, 2009; Yin, 2003). Following Mohr’s (1982) suggestion for process research, this research key focus was on understanding the causal dynamics of a particular setting.

#### 3.1 Research setting

The research setting for this study was the high-velocity internet two sided platform industry, which enables direct transaction or value creation over web-based virtual platforms by linking markets from different groups of users, and extracts a significant proportion of its revenue from such transaction (Zeng and Glaister, 2016). This industry is attractive for this study because data are its core product. Rather than largely relying on physical assets to drive efficiency, the internet platform industry largely depends on their ability to generate information/data – mainly knowledge-based assets that enable/facilitate the interaction between different groups of users in order to create value (Parker and
Van Alstyne, 2005). Consistent with theoretical sampling, we selected companies in which our focal phenomenon of value creation from big data was likely to occur. Specifically, as suggested by Rouse and Daellenbach (1999), we focused on selecting the key performing companies from a single industry to improve the potential for generalizability of our research findings. Following the advice proposed by Block and MacMillan (1985), four companies were selected that were closely matched in terms of starting conditions, availability of resources and company development as factors associated with competitive advantage (Lieberman and Montgomery, 1988) and entrepreneurial growth (Aldrich, 1999; Naman and Slevin, 1993). This research design also enabled the emerging conceptual insights from one case to be evaluated against comparative evidence from the others (Yin, 2003). Table I describes the four cases used in this paper. We stopped at four cases because we were near or at a saturation point and were also reaching the limits of the amount of data that could be processed in one study (Yin, 2003; Pagell and Wu, 2009).

3.2 Data collection

For each company, we traced the process of value creation from big data through both primary and secondary data sources. The primary sources were semi-structured interviews conducted with individual informants. We selected our informants from different departments that were involved in the data analysis and data execution process and from different hierarchical levels, ranging from top management executives to individual data analysts. The main benefit of this approach was that it ensured exposure to different perspectives to compensate for any individual informant personal bias and lack of knowledge, and to enable the cross-checking of the information provided by different informants (Huber and Power, 1985). We employed semi-structured interviews as they afforded us the flexibility to probe informants for details and provide as wide a scope as possible while ensuring that we still covered the issues relevant to our research question (Yin, 2003). The semi-structured interviews were conducted in Chinese, ranged from 60 to 150 min long (but occasionally took as long as 3 h and 30 min), were recorded (if allowed by the interviewees), and were transcribed verbatim within a week by a professional transcribing and translating service provider.

Following Pettigrew’s (1990) suggestion for case-based research, although we approached the organizational field with theoretical constructs in mind, we did not impose them. We carefully considered how the evidence gathered from both primary and secondary data could inform existing theory or constructs, such as resource orchestration and EO. We examined

<table>
<thead>
<tr>
<th>Company</th>
<th>Founding year</th>
<th>Type (from inception)</th>
<th>Ownership</th>
<th>Number of informants</th>
<th>Data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targar</td>
<td>1998</td>
<td>Online social network</td>
<td>Private</td>
<td>9</td>
<td>Reports and strategic memos (25) Press articles (38) Semi-structured interviews</td>
</tr>
<tr>
<td>Yogy</td>
<td>1998</td>
<td>Online group buying site</td>
<td>Private</td>
<td>9</td>
<td>Reports and strategic memos (14) Press articles (19) Semi-structured interviews</td>
</tr>
<tr>
<td>Gray</td>
<td>1999</td>
<td>Online travel agent</td>
<td>Private</td>
<td>7</td>
<td>Reports and strategic memos (11) Press articles (23) Semi-structured interviews</td>
</tr>
</tbody>
</table>

Table I

Background characteristics and data sources for cases
how the data informed our understanding of the process of creating value from big data, and
the mechanisms that drive and facilitate the process.

Our exploratory interviews, in a semi-structured format, were conducted with informants
from the top management executive level of each company as they had “interpretational”
roles (Bennis and Nanus, 1985; Smircich and Morgan, 1982) and “visibility” of the object of
the inquiry (Pettigrew, 1990). The interview protocol involved by asking about the
respondent’s background and the company’s big data strategy. The informant was then
asked to describe the process of value creation from big data and to identify the key
mechanisms that facilitated/hindered this process. After the initial top management
interviews, we conducted semi-structured interviews with staff from various departments,
and then with individual data analysts. The interviews began with a request to describe the
company’s big data strategy and the informant’s personal background. Each informant then
described his/her interaction with the big data team, and the key mechanism that facilitated
or hindered the process of value creation from big data. Thus, a general view of the
mechanisms affecting the process of value creation from big data within the company
emerged. Following the methods of inductive research, these questions were supplemented
with others that seemed fruitful to pursue during the interview. In total, 36 interviews were
conducted. In order to ensure the credibility of the data, we followed the suggestions made
by Eisenhardt and Graebner (2007) and adopted a “courtroom questioning” style, by which
the informants were encouraged to provide concrete examples to support their commentary
and concentrate on facts and events, rather than on their interpretations of them. Complete
anonymity was promised in order to encourage the participants to give candid responses.

Secondary data were also collected to triangulate and gain a complete and accurate
picture (Yin, 2003); these included reports and strategic memos produced by the companies
for the period between February 2008 and March 2013, and extensive archives – including
newspapers, internet sources and corporate materials published between March 2000 and
July 2014.

3.3 Data analysis
As is typical in inductive research, we first built individual case studies using the data
gathered from both the interview transcripts and archival materials. We then wrote a case
study for each site, emphasizing the themes that were supported by the different data
collection methods and confirmed by several informants (Jick, 1979). This was an iterative
process in which we revisited the data as important features of the mechanisms within
each case emerged. We read the cases independently to form our own views of each case
and in order to identify the theoretical constructs, relationships and longitudinal patterns
within each case independently and with respect to our research question. Although we
noted the similarities and differences with other cases, to maintain the independence of the
replication logic (Eisenhardt, 1989), we only started further analyses after we had
completed all the case write-ups.

Once the individual case studies were complete, we conducted a cross-case analysis to
look for similar constructs and themes in the cases (Eisenhardt and Graebner, 2007; Ketokivi
and Choi, 2014; Pagell and Wu, 2009). We started by comparing cases in order to seek
common themes and refine the unique aspects of each particular case. We then used
replication logic to further refine these initial relationships by frequently revisiting each
case in order to compare and contrast the specific constructs, relationships and logics. With
each iteration, we used new permutations of case pairs to refine the conceptual insights. Any
discrepancies and agreements in the emergent theory were noted and investigated further
by revisiting the data. We followed an iterative process of cycling among theory, data and
literature to refine our findings, relate them to existing theories and clarify our
contributions. The propositions were induced following Eisenhardt’s (1989) guidance on
building theory from case studies. After a tentative proposition had been developed, we revisited each case to see whether the data confirmed the proposed relationship. We went back and forth between our data and proposition, relying on the existing literature to further sharpen the insights yielded by the inductive process (Eisenhardt, 1989). We also presented our analysis at a peer workshop and to our informant in order to induce alternative explanations. The feedback we received was taken into consideration when drafting the final conceptual framework. We display additional selected quotes in Table II to illustrate and document the robustness of our claims.

4. Findings and analysis
What emerged from our data were insights that linked value creation from big data with a set of mechanisms. For all companies, making sense of the high volume, high velocity and high variety data itself was central to the challenge of creating value from it. We found that, to address this challenge, companies differed in their approaches to create value from big data.

Through our examination of the data, we developed a framework that capture the value creation process from big data (please see Figure 1).

In the next sections, we elaborate on these insights and describe their grounding in the data.

4.1 Resource coordination for data exploitation
Prior research suggested that the analytical skill and knowledge of data scientists contributes greatly to a company’s opportunities for value creation from big data (e.g. Davenport and Patil, 2012). Yet, our data suggest that the presence of a data scientist or a group of data scientists is a necessary but insufficient condition to cope with high volume, high velocity and high variety data. We found evidence that, while some companies rely heavily on data scientists or data departments to exploit big data, others focus more on bridging the knowledge gap and building coordination networks between the data department and the rest of the company.

Serong provided a compelling illustration of this pattern. Serong had initially set up a data team focusing on data mining. However, the outcome was barely satisfactory, as one informant pointed out:

We have a set of statistics and report from them (data team) on a regular basis. It was useful to a certain extent. They are data analysts, not marketer, not product developers so they could not see much connections and potential as marketers or product developers do.

Following such observation, Serong encouraged the data team to build close collaborative relationships with other departments. Such collaborative interaction stimulated a great flow of knowledge across different departments:

People from different background and discipline see data and correlations from different perspective. For example, data analysts from computer science background would miss or overlooked some correlations and patterns that would matter greatly from marketing and product design perspective. Getting them working together to fully appreciate the meaning from the data is crucial.

Information gathered from other informants and archival data also supported such collaboration.

A similar example can be found from Targar. A consistent pattern across all Targar informants highlighted the importance of broadening the data mining boundaries beyond the individual team/department.

For example, one informant indicated:

Unless you have a clear set of questions in mind when you interrogate data, you can easily get lost given the significant volume of data we have. And because we have so many data, you can also
**Key themes**

**Resource coordination**

"Often we found that people from different disciplines, from different backgrounds often see data from different ways. They also see the pattern in different ways, how it connects to the knowledge they know. When you get people work together in a collaborative way, that's when you really unleash the real potential of the data."

"It's not just about plucking numbers, understand the correlations, this is just a very beginning of the journey, it's about understanding the business context and scenarios. In order to understand the context, we cannot do this by ourselves, we need to talk people from marketing, product development, or even partners from outside of the companies such as local communities to understand the contextual condition."

"Our data team used to be a lone wolf, but now they have people who work on the data, they also have people who are out and about talking to people from different departments, working on different project together. The job we have is not focusing on the data itself, it's about how to make the rest of the people from other departments, make their job much easier."

"We can draw many correlations, but the ultimatum question is always coming down to 'so what'. Data without action, data without context is just data, not more. The action and context part requires collaboration far beyond the data team boundary."

"Everyone jumped on the backbone of this data thing and everyone is fighting to get anyone who is good at it. For me, it is more like 'ji le (chicken ribs)', tasteless when eaten but a pity to throw away. They do daily report and look at the data in the morning, but most of the time, it is more like a procedure, do we get much from it, not really."

"Sometimes people don't really appreciate the job we have done. We worked really hard on the data, but sometimes, nobody even look at our data. They complained that it is not useful."

"I would say the expectation was very high (from us). It is often along these lines: 'tell me how I can steer my company based on the data, or tell me how I can make money based on the data.' By throwing us in the data sea and expect us to know everything is not data management. We felt pressed and isolated."

"We had a bottleneck between the data analyses and data application. How to breakdown the data and feed into different department. They (data analysts) know how to fiddle around data, how to make sense the data, but the hidden layer is that they need to understand the business value of these data. They present the data on the daily basis to us, but most of us, to be honest, do not really know what we look at, what these data means."

"Data itself, particularly the real time data is very valuable. You have to be prepared, to anticipate the changes, and you have to act fast. If you are not prepared, by the time you put everything together and act on it, the opportunities are gone. You want our front line staff to work with the data."

"Somehow we feel that we always play a catch up game. With layers after layers of management, with layers after layers approval, we were never been capture the real value of big data. The real value is not about mining the data we have in the past, this is just too passive and you won't..."
being generous. This means that in order to capture this opportunity, simply responding is not sufficient, we need to take action drive the change and anticipate the change.”

“With amount of the data we have, we are in a better position to make decisions about the future. We know what to expect and how to approach it.”

“We often neglect the human side of the data business and how people can use data in a different and innovative ways. We tried to strive for a balance between analytical and innovative side of data.”

“The term ‘big data’ sounds intimidating and very technical. We wanted to change this perception and encourage our staff to have fun with it. That’s why I say data cannot overtake all the jobs because there are certain aspects such as curiosity, creativity and imagination, things we are good at but data cannot do. That’s exactly what we try to get our people to do when they approach the data, be curious, ask questions, be creative, think in different ways.”

“People got fascinated about data itself and forgot important mechanism to enabling to make these data ‘alive,’ the mechanism I am talking about is culture. You need to have that culture change, the culture that takes an innovative approach when it comes to data.”

“It is important to create a synergy between data and people. Data is important proving solid evidence, but it can only be to its best potential with the magic touch from people.”

“Everything is ‘we need to look at the data’ or ‘let the data do the talk’. We don’t have the voice in the process. When it comes to data, we often have a blurred vision, it is like to admire flower in fog weather, and nobody really know the true beauty of the flower.”

“The emphasis is always on the analytical side of data. Innovation, well, they (data team) just have to think something different.”

“Crunching big data requires hard-analytical skills. We have great experts in the house.”

“We put great investment into our data team, focussing on the data analytical side of business.”

---

**Table II. Value creation through big data**

<table>
<thead>
<tr>
<th>Key themes</th>
<th>Serong</th>
<th>Targar</th>
<th>Yogya</th>
<th>Gray</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Innovation</strong></td>
<td>“We often neglect the human side of the data business and how people can use data in a different and innovative ways. We tried to strive for a balance between analytical and innovative side of data.”</td>
<td>“The term ‘big data’ sounds intimidating and very technical. We wanted to change this perception and encourage our staff to have fun with it. That’s why I say data cannot overtake all the jobs because there are certain aspects such as curiosity, creativity and imagination, things we are good at but data cannot do. That’s exactly what we try to get our people to do when they approach the data, be curious, ask questions, be creative, think in different ways.”</td>
<td>“People got fascinated about data itself and forgot important mechanism to enabling to make these data ‘alive,’ the mechanism I am talking about is culture. You need to have that culture change, the culture that takes an innovative approach when it comes to data.”</td>
<td>“It is important to create a synergy between data and people. Data is important proving solid evidence, but it can only be to its best potential with the magic touch from people.”</td>
</tr>
<tr>
<td>Key themes</td>
<td>Serong</td>
<td>Targar</td>
<td>Yogya</td>
<td>Gray</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>creative and use their imaginations, use data to create different stories, to create emotional connections with our customers”</td>
<td>“It is great that data can provide many insights about our customers’ behaviour but if everything is driven from data, you never get opportunities to take risk and try it out new things. People often afraid to take a risk, but you have to be all in” “We have the luxury to try it out in a small scale to see how market response. Without taking risks, you will be blinded by many other potential opportunities in the market”</td>
<td>“Data is a very important resource, but data is a history, we need to discover, create new opportunities, not just from the data we have, but to create new insights, to shape the data. This requires us to take a plunge, yes, it can be risky, but if are afraid of making mistake” “If you are afraid of taking risk, then you are definitely in the wrong side of big data business. It is not about play safe with the data you have, it also about explore new ideas and opportunities, to create and drive the future with the help and insight from big data”</td>
<td>“You don’t have to take risk because the insights generated from big data can help you to unravel the misty of the phenomena, so why do we need to try something different” “The risk is greatly reduced with the power of big data. It provide us with a much safer environment to do our business”</td>
<td></td>
</tr>
<tr>
<td>Risk taking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
easily get correlations that make no sense at all. So we talked to people from different departments, to understand what is bothering them that can be solved by the data we have, what data or information do they need to make their job better and easier.

This view was often echoed by other informants, who ascribed great significance to cross-department collaboration. We also noted that Targar had opened up its platform to external partners to develop value-added applications based on the data sets available on their platforms. A senior director from Targar commented:

We were sitting on huge amount of data that was not being used to its full potential. We only have limited resources here and by opening up our dataset on our platform really connect our data with many ‘a-ha’ moment. They [the external partners] never failed to impress us with their creative ideas.

Similarly, many informants highlighted that opening up their data to third-party developers had resulted in the establishment of a distinctive resource network within their network innovation systems.

We observed quite different patterns in the other cases. Yogy, for instance, did not mention such cross-disciplinary collaboration. Yogy informants expressed rather frustrating views on benefiting from big data:

They always sit behind closed doors and we never know what is going on there. We do regularly get reports and instructions from the top [the managers] in terms of what to do based on the data they [the data team] crunched but sometimes it just seemed rather pointless and a waste of time. Yes, they are good at numbers, but do they understand business, or do they understand our jobs?

The findings suggest that the managers from Yogy had often had high expectations in terms of the data department generating value from big data. One informant from the data department explained:

We are under a lot of pressure to deliver results, but our attention is limited given how much data we have crunch. And data crunching is not just a one-off, it changes especially given how many real time data we have. It should be about the instant flow and connection with other teams. We feel quite isolated here.

Similar patterns were observed across another case company – Gray. The data team was often described as “living in their own world.” The informants often expressed a low level of confidence in the role of big data holding strategic importance for their companies. One informant commented:

It is a big hype. We implement changes that were requested from the top from the analysis reported put together by the data team, some of the requests did not make much sense and we never got to fully understand it. I wish a bilateral dialog could be built.

Why does resource coordination between different departments facilitate the process of value creation from big data? One reason may be that those who lack familiarity with customers (Shane, 2000; von Hippel, 1988) and knowledge of ways to serve the market (Shane, 2000) find it difficult to recognize solutions to meet customer needs and to formulate

![Figure 1. Framework](image-url)
effective strategies to introduce and sell new products/services. Therefore, by working in isolation, data scientists are unable to discover patterns and correlations related to customer behavior. This was illustrated by the following quote:

Given the volume of the data we have, it is easy to get random correlations. We need people with different backgrounds and knowledge of the market to truly appreciate and understand these correlations and patterns. You cannot achieve this by getting them to work separately.

Based on our findings, without such coordination, an organization is less capable of discovering and exploiting new opportunities emerging from big data. Many scholars highlighted that team members’ willingness and ability to share hard-to-find specialized knowledge with other team members (bringing expertise to bear) were crucial to firm’s performance (Majchrzak et al., 2000; Kanawattanachai and Yoo, 2007). From the above evidence, we argue that resource coordination as a result of cross-department collaboration provides a company with an increased ability to discover and exploit any opportunities emerging from big data. This leads to our first proposition:

\[ P1. \] Resource coordination between different departments is positively related to opportunities for discovery and exploitation from big data which leads to value creation.

4.2 Entrepreneurial orientation and value creation from big data

The findings indicate that our sample companies varied in their approaches to create value from big data. Some believed that employing skilled data scientists was critical to success. Others valued data scientists but ascribed more importance to capitalizing on big data by encouraging entrepreneurial activities at the company level (Covin and Lumpkin, 2011).

One good example was provided by Serong. The main focus of Serong was to encourage its employees to be curious about data and to experiment with it. One informant had this to say about his understanding of the key aspects of creating value from big data:

We wanted people to take ownership of the data, not to think that data analysis is the experts’ job and responsibilities. They don’t need to get to the technical side of it, but rather to think creatively about how we can tell a good story with the data we have, what kind of impact it will have on our customers.

Creativity and experimentation were perceived as great approaches to complement the technical data crunching. For example, Serong introduced a “Magic Data Cube” initiative to inspire employees at the company level to generate projects based on the big data. One informant from data team described:

Data are like a magnifying glass to understand the market, its trends and our customers. There were many interactions between us [the data department] and the rest of the company; what data they needed to tell a story and what data we had to support that. They can see things differently and come up with great creative ideas which we would never have thought of.

Similar entrepreneurial examples can be found from Targar. Many informants highlighted the value of real time data and the importance of a company’s ability to extract information from real time operational data. The faster a company can harness insights from real time data, the greater its advantage in driving its value creation opportunities, as “top down command-control structure will not work well in this sense.”

One informant commented:

By the time you get through the different layers of approvals, you already missed the time and opportunity to respond to these data. The top down structure has to change to cultivate and support the entrepreneurial activities at the front line level.
Our informants further accentuated a risk-taking approach in relation to data management. This was explained by the following observation:

With the insights coming from the data, you don't have to go all out. You can manipulate the scale to which you want to test your ideas. Initially, the scale is quite small, so the risk is low; then, you can gradually scale it up based on how data reacts.

Such entrepreneurial activities were perceived as being essential in driving a company's value creation opportunities from big data.

These findings support the notion of EO in maximizing value from big data. The data suggest that the companies with a greater level of EO were in a better position to take advantage of the opportunities offered by big data. The literature has consistently identified three dimensions of an EO: innovativeness in engaging in creativity and experimentation, thereby departing from established practices and technologies (Lumpkin and Dess, 1996); proactiveness in opportunity-seeking, being forward-looking to stay ahead of the competition (Lumpkin and Dess, 1996); risk-taking, being willing to commit large amounts of resources to projects in which the cost of failure could be high or the outcomes unknown (Miller and Friesen, 1978). The underlying assumption is that an EO provides organizations with a basis for entrepreneurial decisions and actions for capturing innovative opportunities (e.g. Lumpkin and Dess, 1996; Wiklund and Shepherd, 2003; Covin and Lumpkin, 2011).

By contrast, such entrepreneurial actions were barely mentioned by Gray and Yogy. For example, Gray emphasized the technical side of big data and the importance of the data department in contributing to a company's value creation opportunities from big data. When asked about the involvement of other departments in value creation opportunities from big data, one informant from the top management team described:

Of course, they are given the data analysis report – sometimes from us, sometimes directly from the data team. It is a chain of action and they can act on what needs to be done.

When asked the same question, an informant from a different department stated:

We do receive regular reports and tasks from them [the top management and the data team], some of them are useful but most of them are not contextual, there is no story behind it. I wish that we could get more involved in this process.

We observed very similar patterns in Yogy. The Yogy informants described the process of generating insights from big data as being too technical; therefore, the involvement from other teams/departments was limited. Some of the interviewees stated:

We just follow the direction and act on what needs to be done.

When asked about experimentation and risk taking, an informant responded:

They focus too much on the left brain, it is all about analytical and logic. Data dictate everything and we play limited roles in the process.

Our analysis points to the key insight emerging from the above evidence: that those organizations that have an EO are more likely to generate value from big data. An EO represents the way a company is organized in terms of utilizing resources in order to uncover and exploit opportunities. Based on the RBV, how a company is organized, when coupled with its resources, can increase the positive relationship between resources and company performance (Barney, 1995). The findings demonstrate that an EO captures the way a company is organized toward entrepreneurship and can enhance value creation opportunities from big data. Our data reveal that the key elements of an EO – such as innovativeness, risk taking and proactiveness – can partially explain the process of value creation from big data that enables some companies to get ahead of their competitors.
Companies with higher levels of EO can be in a better position to effectively utilize big data-related knowledge assets for both the discovery and exploitation of opportunities arising from big data analytics (Covin and Lumpkin, 2011; Davenport and Patil, 2012; Khan and Vorley, 2017). Previously, we proposed a positive relationship between resource coordination between different departments and value creation from big data. Because of the magnitude of the data volume and of the speed at which it should be analyzed and acted upon, we further propose that a managerial decision-making process that champions a willingness to capitalize on its big data resources by engaging in entrepreneurial activities at the company level will perform even better in creating value from big data (Wales et al., 2013). This leads to the following proposition:

**P2.** An EO moderates the relationship between resource coordination and value creation from big data.

### 5. Discussion and conclusions

The aim of this paper was to understand the value creation through big data in dynamic environments, such as those observed in emerging economies. Recently, there has been an increasing interest in understanding the role of big data and its resultant implications for performance and knowledge management (Jabbour et al., 2017; Dubey, Gunasekaran and Childe, 2018; Khan and Vorley, 2017). The existing studies have enhanced our understanding on this topic, yet the research on big data is at its infancy and further research has been suggested in developing solid understanding about how firms co-create knowledge and capture value through big data (e.g. Acharya et al., 2018; Dubey, Gunasekaran and Childe, 2018; Wamba et al., 2017; Khan and Vorley, 2017). In order to understand value creation through big data in emerging economies, we integrated resource orchestration (e.g. Sirmon et al., 2007; Sirmon et al., 2011) and EO (Lumpkin and Dess, 1996; Covin and Lumpkin, 2011).

The findings highlight that resource coordination is vital for firms to create value through big data by firms based in emerging economies. The findings further indicate the important role of EO through which resource coordination and asset orchestration lead to the value creation from big data in emerging economies.

#### 5.1 Theoretical contributions

Our study offers several insights for business management in the big data context (e.g. Acharya et al., 2018; Akter et al., 2016; Wamba et al., 2017). The primary contribution of this study is that we explicitly incorporate resource orchestration into the domain of big data and identify EO as one of the key factors through which companies bundle and orchestrate the knowledge assets arising from big data for value creation. While the three frameworks of resource management, asset orchestration and resource orchestration have been established to describe the use of resources to create value (Helfat et al., 2007; Sirmon et al., 2007; Sirmon et al., 2011), additional empirical research is needed in order to add richness to current theory (Sirmon et al., 2011; Chadwick et al., 2015). The findings indicate that resource coordination between different departments facilitates the process of value creation from big data. We have, thus, extended the understanding of resource orchestration in a big data context, whereas previous research has provided key insights by utilizing dynamic capabilities approach (Wamba et al., 2017). Our findings are consistent with the hitherto largely underexplored arguments that resources themselves may not create value for companies; it is how companies utilize those resources that is important in explaining corporate performance. The findings further indicate that an EO moderates the relationship between resource coordination and value creation from big data. That is, the willingness to be innovative, proactive and taking risks enhances a company’s capability to generate value.
from big data (Covin and Lumpkin, 2011). These findings echoed Chirico et al.’s (2011) observations and suggest that an EO can help explain the managerial processes that provide some companies with the ability to utilize their resources to identify and respond to environmental cues earlier than competitors.

Second, while most of the existing literature on resource orchestration is conceptual in nature (e.g. Sirmon et al., 2007, 2011), this study provides important empirical insights demonstrating the value of integrating resource orchestration and EO in explaining the performance implications of big data. Our findings are, therefore, consistent with the existing resource orchestration and dynamic capability conceptual apparatus, such as evolutionary and entrepreneurial fitness (Helfat et al., 2007). The findings of this study provide an important foundation to explore how resource orchestration and EO may influence the process of generating value from big data. Our research, thus, could be adopted in further studies as a starting point from which to examine the effectiveness of an EO in shaping resource orchestration to enhance value creation from big data.

In addition, we provide important insights from the important emerging economy of China and show how managers bundle and orchestrate resources and create value from big data. This is one of the first studies to have integrated resource orchestration theory and EO in the context of big data, thus demonstrating the value of integrating the RBV with EO in the context of big data. Not much is known about how companies from emerging economies orchestrate and leverage knowledge assets, particularly the valuable knowledge that can be captured through big data for value creation strategies; therefore, the findings of this study have greater value for managers in order to understand the creation of value from big data through the adoption of specific managerial processes and strategies that are conducive to the orchestration and leveraging of resources for the development of competitive advantage in the era of big data (Prescott, 2014).

5.2 Implications for practice

The findings of this study provide important implications for practice. The findings suggest that resource coordination is important in harnessing value from big data in dynamic environments of emerging markets. Thus, managers need to carefully orchestrate asset base and coordinate internal resources in order to benefit from big data for value creation which will lead to the development of competitive advantage (Barney, 1991; Helfat et al., 2007; Sirmon et al., 2011, 2007). The resource coordination for exploitation of big data can be further realized through proactive, innovative and risk-taking EO, therefore, managers need to improve and reconfigure internal processes in order to create value through big data.

5.3 Limitations and further research

Our aim was to gain a rich understanding of how companies manage big data resources to create value in the context of emerging economies. With regard to generalizability, it is critical to note that the sample companies operated in a data-intensive area and had very different starting points. However, with companies in other industries generating more data than ever before, the process of value creation from big data that we observed may be generalized to other ventures. We also acknowledge that this research’s setting was limited to China. Therefore, future research can expand or test (e.g. by using quantitative methods) across industries and countries the two propositions that we have put forward. Additionally, future studies could examine micro-processes and other strategic orientations – such as learning and marketing orientation – and how these influence the creation of value from big data.

Summarizing, one of the key contributions of this study is that it brings the EO into the discourse on resource orchestration and value creation from big data. A key implication of our findings for resource orchestration scholars is that the investigation of the relationship
between big data and value creation should also consider its organization. Therefore, further research needs to investigate the effectiveness of an EO in affecting the explorations of ways in which a company is organized for the detection and exploitation of opportunities from big data. In addition, there is a need for more scholarly attention to the development of company control and coordination processes aimed at encouraging and stimulating EO behaviors in individual employees, thus, focusing on the central role played by persons and interpersonal interactions in harnessing EO – as pointed out by Salvato and Vassolo (2018) – rather than that of abstract, company-level entities. Lastly, the insights of this study can be applied in the context of sharing economy-based firms in order to understand those firms value creation strategies. Furthermore, there is a scope to examine other antecedents such as learning orientation, social networks and absorptive capacity and how these enable firms to create and capture value from big data across emerging and developed markets.

References


**Further reading**


**Corresponding author**

Zaheer Khan can be contacted at: Z.A.Khan@kent.ac.uk

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com
Influence of basic research investment on corporate performance
Exploring the moderating effect of human capital structure
Malin Song
Research Center of Statistics for Management, Anhui University of Finance and Economics, Bengbu, China, and
Xiongfeng Pan, Xianyou Pan and Zhiming Jiao
Faculty of Management and Economics, Dalian University of Technology, Dalian, China

Abstract
Purpose – The purpose of this paper is to add to the existing research about how corporate performance is influenced by their basic research (BR) investment. On this basis, the authors examined the moderating effect of human capital structure (HCS) on the relationship between BR investment and corporate performance.
Design/methodology/approach – Hypotheses were tested using static and dynamic models to analyze a large-scale data of Chinese A-share listed companies.
Findings – This study provides empirical evidence that contributes to the research about how private BR investment influences corporate performance in the digital age. In addition, human resource is an important dynamic ability for enterprise development. Based on the dynamic capability theory, further research finds that the human resources practice on the knowledge stock can enhance the company’s dynamic capability, thereby enhancing the company’s core competitiveness.
Research limitations/implications – The results may be affected by the context of the data set. This study considers the influence of private research investment type on corporate performance, further studies considering the influence of specific contextual variables, such as corporate industry differences, could yield richer insights that would help validate the results of this study.
Practical implications – This study provides useful information for managers. As well as increasing the investment in the BR of enterprise and creating the necessary conditions to increase the competitiveness of enterprise, they should strive to adjust the structure and quality of researchers involved in BR projects.
Originality/value – This research contributes to the enterprise's BR investment and the management of human capital resource. It points that the investment of BR positively influences the corporate performance. In addition, the increasing of high-skilled labor’s proportion positively promotes the promotion of BR investment on corporate performance.

Keywords Moderating effect, Corporate performance, Basic research investment, Human capital structure

1. Introduction
The Chinese economy has experienced a long period of rapid growth, and it has created the so-called “China Miracle” (Tong et al., 2008; Tian, 2016; Tseng and Bui, 2017). However, since the financial crisis occurred in 2008, the pressure of economic downswing in Mainland China has gradually increased, and the traditional economic growth model driven by investment and export has been faced with a tremendous challenge. Taking the innovation-driven development as a major national strategy is an urgent task for China during the period of comprehensive economic and social transformation (Choi et al., 2011; Xie et al., 2015; Shao et al., 2016; Sesay et al., 2018). Especially in the context of supply-side reform, enterprises in China should innovate to drive the increase in total factor productivity (TFP) and achieve greater efficiency (Aarstad et al., 2016). The investment in research and development (R&D)
of enterprises is an important factor in increasing the total amount of knowledge and then using new knowledge to create new applications (Mendigorri, 2016; Yu et al., 2016; Jabbour et al., 2018). The scale and intensity of investment activities often reflect the scientific and technological strength and core competitiveness of regions and enterprises (Su et al., 2018; Wang et al., 2018).

In this paper, we elaborated on the impact mechanism of enterprise R&D investment on corporate performance at first. On this basis, taking into account the importance of basic research (BR) investment in creating advanced technology and opportunities for enterprises, this paper subdivided enterprise R&D investment and researched on the relationship between enterprise BR investment and corporate performance. Second, due to the differences in the conditions of enterprises, there are significant differences in the impact of BR investment on corporate performance among different enterprises. Based on the perspective of enterprise dynamic capabilities, this paper theoretically expounded that human capital as an important strategic resource of the company, which has a significant impact on the company’s strategic decision-making and operational processes. Further, we constructed a reasonable empirical analysis model, which deeply confirmed the moderating effect of human capital structure (HCS) on the relationship between BR investment and corporate performance, thus providing theoretical and empirical support for better playing the role of BR investment, and also providing inspiration for corporate managers to make reasonable and effective R&D investment decisions.

The remainder of this paper is organized as follows: Section 2 gives a comprehensive review of the relevant literature. Section 3 builds the hypotheses according to theoretical analysis. Section 4 presents the research design and data processing. Section 5 presents empirically analysis and results. In Section 6, conclusions and implications of this research are covered, and limitations and certain directions for future research are indicated.

2. Literature review

Up till date, how to evaluate the economic value of R&D activities of enterprises has become an important issue. It has been widely concerned by scholars at home and abroad for a long time. Theoretically, both in innovation economic theory and new growth theory, R&D investment was crucial among the fundamental drivers of enterprise (Lucas and Ayse, 2018). First, private R&D investment enables enterprises to create new products, develops efficiency processes, and stimulates strategic cooperation (De and Freel, 2010). Second, taking the long view, it brings the development opportunity for the enterprise, thus affecting the enterprise’s survival ability and increasing the competitive advantage of enterprise (Cohen and Levinthal, 1989). Sougiannis (1994) pointed out that the increase of R&D investment has a significant positive effect on the improvement of corporate performance. Singh (2009) utilized a two-stage least square estimation on a sample of 47,140 firm-year observations over a period of 16 years from 1990 to 2005, and pointed that R&D expenditure positively affects enterprise’s export sales. Bosworth and Rogers (2010) confirmed that there was a significant positive correlation between R&D investment and corporate performance based on the Tobin’s Q-value method. From the perspective of R&D investment intensity, Falk (2012) believed that it has a significant positive impact on sales revenue growth, thus effectively boosting corporate performance. Gaur et al. (2014) integrated the resource- and institution-based views and examined in the broader context of emerging economies. They found that firms that have more technological resources are more likely to gain benefits and advantages. Park and Ryu (2015) pointed out that R&D capability and learning capability were significant drivers of SEMs technology commercialization, which in turn influenced their business performance. Meanwhile, environmental dynamism was found to moderate the relationship between technology commercialization and business outcomes. Based on the data collected from 560 service and manufacturing companies in Germany, Pippel and
Seefeld (2015) considered that enterprises with closer R&D cooperation with university and government have a greater willingness to invest in R&D, which promotes an obvious increase in product innovation capabilities and has a positive impact on corporate performance. Boles and Link (2016) selected typical start-up companies in ten European Union countries, and found that enterprise’s sales increase rapidly when R&D practice becomes the main investment, and ultimately corporate performance will be significantly improved. Yeom (2017) estimated the effect of R&D investment on technological innovation of ICT SMEs using binary logistic regression analysis based on 2016 ETRI Survey. The authors pointed out that R&D investment shows a positive effect on both product and process innovation. Similarly, Sun (2018) pointed out that R&D intensity was found to be positively significantly related to corporate performance.

Contrary to the views mentioned above, past scholars believed that R&D investment has a negative impact on corporate performance. Private R&D activities result in huge operation costs but incur low diffusion cost, that is why enterprises need to maintain certain technological frontiers and increase limitation difficulty, and ensure that their R&D activities are profitable and prevent other companies to benefit from R&D activities achievements (Yeh et al., 2010). Thus, for some individual companies, restricted by many factors, such as the company’s development stage, scale, etc., the smaller scale of R&D investment and the lower utilization efficiency of R&D fund can hardly promote corporate performance. Second, R&D investment has an opportunity cost effect, with the increasing of R&D invest scale, it will inevitably occupy the input of other factors and reduce the enterprise’s profitability (Yang et al., 2010). Finally, most of R&D expenditures are used to pay R&D personnel’s salary, and the revenue created in the future for the enterprises is an intangible asset. Limited by the long R&D investment cycle, any disruption in the course of R&D activities may result in the loss of R&D personnel and corporate performance (Ju et al., 2013). Nunes et al. (2012) argued that there was a negative linear relationship between R&D intensity and growth in non-high-tech SMEs, because R&D investment can increase the level of risk faced by SMEs and this risk increases the difficulty of obtaining financial support. Finally, some scholars have also argued that enterprises, limited by the firm scale, tend not to have the means to invest in private R&D activities (Avermaete et al., 2003) or cannot always transform R&D into effective innovation (Laforet, 2009), which has an inhibitory effect on the improvement of corporate performance.

Finally, with the continuous innovation and development of information technology and the rapid and deep integration of the internet and the real economy (Tseng, 2017), private R&D expenditure management has entered the era of big data (Blackburn et al., 2017; Holden, 2016). Considering the increasingly obvious internal remodeling pressure and the uncertainty of the external environment, how to obtain a lasting competitive advantage in this highly dynamic environment has become a research focus in the field of enterprise innovation and performance management (Wu, 2006; Singh and Gaur, 2013; Kim et al., 2015; Shao et al., 2017; Singh et al., 2018). In fact, enterprises used big data, gathered from social media steams, sensors embedded in consumer products and elsewhere to identify problems and improve ideas with a newly launched research project (Sharwood, 2015; Singh, 2018). In addition, dynamic capability, as a special ability, is the same as the organization’s other capabilities and is embedded in the organization (Salvato and Vassolo, 2018). More important, the dynamic capability exists in the form of an “intermediary,” with the acquisition of resources as a prerequisite, focusing on improving the efficiency of enterprise resource utilization, and ultimately helping enterprises to obtain economic profits and competitive advantage (Rodney et al., 2011; Teece et al., 1997; Dubey, Gunasekaran and Deshpande, 2017; Dubey, Gunasekaran, Helo et al., 2017; Nuruzzaman et al., 2017). This also means that the resources of the enterprise do not directly affect the enterprise’s competitive advantage in the dynamic environment, but influence enterprise’s competitive advantage
through dynamic capability. Human resource is an important support element for enterprise development; the link between human resource and corporate performance has received significant research attention (Ogunyomi and Bruning, 2016; Young, 2017; Chung, 2010). All of the scholars believed that human resources management has a positive impact on corporate performance (Hazen et al., 2017; Youndt et al., 1996; Hitt et al., 2001). In this paper, combined with the dynamic capability theory, the authors believed that the human resources practice on the knowledge stock can enhance the company’s dynamic capability, thereby enhancing the company’s core competitiveness in the digital age.

Based on the review of the above articles, R&D activity is one of the effective ways for companies to improve their performance, and the relationship between them has attracted the attention of many scholars. However, the confusion of empirical conclusions still brings great uncertainty to the timing of government and business decisions. In addition, the authors often used private R&D expenditure as a corporate research investment in the previous study, which includes basic R&D application. Among them, BR is not only the source of explicit information, but also creates new technological opportunities and sustainable development opportunities (Klevorick et al., 1995; Salter and Martin, 2001). Thus, the research focuses on the original and long-term mechanism of BR expenditure, which has a significant relationship with corporate economic development. At last, human resource is the core element of the dynamic capability of the company in the digital age, but few literatures have focused on the moderating effect of this factor on BR investment.

Our contributions to the corporate performance literature are twofold. First, we have further subdivided enterprise’s R&D expenditure, studied the relationship between BR investment and corporate performance from theoretical and empirical two perspectives, and supplement and expand existing research. Second, we emphasized the significance of HCS in accentuating the implications of BR investment for corporate performance. We have demonstrated that an enterprise’s degree of HCS can enable it to better exploit BR investment for higher corporate performance.

3. Theory development and hypotheses
3.1 BR investment and corporate performance
From the perspective of short term, BR requires a large amount of capital investment. Limited by capitalization conditions, BR investment costs will inevitably deteriorate the corporate financial status and reduce corporate performance (Yang et al., 2010; Nunes et al., 2012). However, from the perspective of long term, BR investment will promote the improvement of corporate performance. First, through organizational learning and knowledge sharing in the BR investment process, the accumulation of heterogeneous knowledge can enhance the absorptive capacity of the company and increase the synergy between the internal and external knowledge technology, which brings improvements to the corporate performance (Lane and Lubatkin, 1998; Levi et al., 2013; Gaur et al., 2013). Especially with the rapid development of science and technology, production technologies are characterized by their integration, complexity, and variability. Companies investing in BR can not only solve problems and create new knowledge, but also improve their technological absorptive capacity (Wesley and Daniel, 1990). The combination of externally acquired knowledge technology and internal existing knowledge and technology can change the company’s existing knowledge stock and knowledge structure (Lee, 2001), resulting in scale effect and knowledge reorganization effect. Second, intangible assets such as patents and non-patent technologies that may be generated after BR investment are also heterogeneous resources of enterprises. According to the resource-based theory, these heterogeneous intangible assets are valuable, scarce, and difficult to imitate and substitute. It can bring sustainable competitive advantages to the company (Barney, 1991;
H1. BR has a positive effect on the improvement of corporate performance.

3.2 The moderating effect of human capital structure on the relationship between BR investment and corporate performance

The BR investment process of a company is characterized by systematicness, complexity, and variability (Lee et al., 2017). The pre-investment strategic decision-making and resource integration in the BR investment process and the commercial operation of new research directly affect the corporate performance (Cereola et al., 2012; Escriba et al., 2009). This also means that the BR investment is a process of continuous strategic decision-making, organization and coordination, and resource integration. It is the continuous operation and application process of dynamic capabilities of enterprises (Teece et al., 1997). These dynamic capabilities include strategic decision-making capabilities, organizational coordination capabilities, and resource integration capabilities (Salvato and Vassolo, 2018).

Hambrick and Mason (1984) believed that the human capital is the most important strategic resource of the company and significantly influences the company’s strategic decision-making and operation process. The human capital includes the education level, professional background, work experience, and social status. This paper draws on the board of directors’ capital model proposed by Haynes and Hillman (2010), and describes the human capital from the perspective of structure. Specifically, the HCS refers to the overall educational level, work experience, and the degree of embedding in the enterprise (Birasnav and Rangnekar, 2010; Jabbour et al., 2017). It also reflects the level of education and professional experience of the research members has an important impact on the transformation of BR investment into corporate performance. First of all, BR investment is an important strategic decision for a company. Prior to making decisions, it is necessary to conduct comprehensive data collection and feasibility analysis on the external environment, internal technical conditions and resource allocation, and future market prospects faced by the company (Singh, 2018; Yun, 2013). As the external environment and future market prospects are extremely uncertain, and internal technical conditions and resource allocations are also in a dynamic change, the cognitive ability, market insight, and values of the enterprise’s human capital directly affect the strategic decision-making process and ultimately affect the quality of strategic decisions of BR (Guo et al., 2018). The level of education and professional experience is higher, the ability of technology forecasting, comprehensive consideration before making decisions, and investment conversion rate is higher, so that they can make correct strategic decisions and help to develop BR investment projects successfully. Second, some authors have studied the influencing factors of R&D team’s innovation performance from the perspective of team organization and collaboration, and believed that the team members’ spirit, common values and beliefs, mutual cooperation, and support have a significant positive effect on team performance (Thamhain, 2003; Al-Ali et al., 2017; Dubey, Gunasekaran and Deshpande, 2017; Dubey, Gunasekaran, Helo et al., 2017). In fact, the members of the BR team are different from the general human resources. They have high academic qualifications and technical background and represent common values and beliefs. Thus, it is easy to further increase the conversion rate of BR investment (Cable and Judge, 1997; Kristof et al., 2005). Third, the better HCS also reflects the large proportion of research member with background in the industry, such as having more related technology and research talents and strategic management talents. These talents have a sense of market crisis, innovative awareness, and stronger predictability for the uncertainty in the BR process (Sutcliffe and Zaheer, 1998), so they can make corresponding
adjustments in time to improve the basic research performance. From this, we propose the second research hypothesis of this paper:

\[ H2. \text{ The HCS positively regulates the relationship between BR investment and corporate performance.} \]

4. Method

4.1 Model design

This study analyzes the influence of BR on corporate performance based on the Mansfield model proposed by Mansfield (1980) and the double log-linear model. The special model is as shown in the follows:

\[
\ln TFP_{it} = a_0 + a_1 \ln BR_{it} + \mu_i + \epsilon_{it}.
\]

In an open business environment, corporate performance not only depends on the BR investment but also on other economic factors. Based on this proposition, this paper added other factors that reflect the company’s basic conditions into the above model as control variables, including enterprise scale, enterprise age, and capital structure (CS). The basic model can be eventually set as:

\[
\ln TFP_{it} = a_0 + a_1 \ln BR_{it} + a_2 \ln X_{it} + \mu_i + \epsilon_{it},
\]

where \( TFP_{it} \) is the corporate performance of \( i \) enterprise in year \( t \). \( BR_{it} \) represents the basic research investment. \( u_i \) is enterprise unobserved individual effect, capturing enterprise \( i \)'s all time-invariant characteristics; \( \epsilon_{it} \) is the error term, we allow for cross-section by clustering the standard errors. \( X_{it} \) represents the control variables.

In order to examine the moderating effect of HCS on the relationship between BR investment and corporate performance, this paper adds the interaction term of HCS and BR in the basic regression model. The basic model can be extended as follows:

\[
\ln TFP_{it} = a_0 + a_1 \ln BR_{it} + a_2 \ln X_{it} + b_1 \ln BR_{it} \times \ln HCS_{it} + \mu_i + \epsilon_{it},
\]

where \( HCS_{it} \) represents HCS of \( i \) enterprise in year \( t \).

4.2 Data collection and sample

In 2006, the Ministry of Finance of the People’s Republic of China stipulated the accounting standards for enterprise intangible assets, subdivided the expenditures of internal R&D projects of the company, and clearly pointed out that there are significant differences in research objectives, nature, and method between basic R&D research, which are required to perform accounting separately. Taking into account the information content of big data, the authors in this paper constructed a large-scale data concludes 3,459 A-share listed companies from 2011 to 2016. All of the data are mainly selected from Resset Financial database and Wind database. In order to ensure the completeness of the data and the matchability of all variables, based on the encoding of the company’s name, the authors used Stata software to match each variable and exclude the enterprises with missing data. The final effective data contain 83 A-share listed companies. The price data are deflated according to the corresponding price index. All of the subsequent empirical analysis in this paper is mainly based on Stata 13.0. Descriptive statistics of each variable are shown in Table I.

4.2.1 Explained variable. TFP with reference to the previous literature, TFP reflects the efficiency of corporate production activities in a certain period (Griliches and Mairesse, 1991; Hall and Jones, 1999) and can be used as an indicator to measure scientific and technological
progress and financial status of enterprises; thus, this paper mainly measures corporate performance with TFP (Song, 2015; Rai and Mahapatra, 2009). From the perspective of economic growth, TFP is equal to the residual value of total output after removing inputs such as labor and capital. In this paper, the author measures the TFP based on the method of Sequential Malmquist Index (Caves et al., 1982).

Measuring TFP requires explicit input and output data. In this paper, the input data mainly include labor and capital and the output variable is the main business income of the enterprise; where, the labor is measured by the total number of employees in the enterprise. The capital stock $K_{it}$ is calculated using the perpetual inventory method, as shown in the following equation:

$$K_{it} = (1-\delta)K_{it-1} + I_{it}.$$  \hspace{1cm} (4)

The capital stock in the base year 2011 is calculated according to the following equation:

$$K_{2010} = I_{2010}/(g_i + \delta),$$  \hspace{1cm} (5)

where $I_{it}$ is the capital investment amount of enterprise $i$ in the year $t$, $g_i$ is the average growth rate of fixed capital investment, $\delta$ is the depreciation rate of capital stock and is equal to 9.6 percent (Zhang et al., 2007).

### 4.2.2 Core variable.
BR: consistent with the investment in fixed assets, the flow data were converted into stock data for the amount of BR investment. It needs to be noticed that the depreciation rate of BR investment is 15 percent in this paper (Jaffe, 1986).

### 4.2.3 Regulated variable.
HCS: according to the research of Asuyama (2012), this paper uses the proportion of highly skilled labor in the total labor to measure the HCS. With regard to the definition of high-skilled labor, this paper classifies high-skilled labor and low-skilled labor according to the International Education Standard Classification issued by UNESCO. Among them, the high-skilled labor is the labor who has received college education and above.

### 4.2.4 Control variable.
4.2.4.1 Enterprise size (Size). The famous “Schumpeter Hypothesis” holds that large enterprises have stronger innovation capabilities than small firms because of their relative advantage in economic size, risk sharing, and financing channels. Meanwhile, a large number of empirical findings also confirm the impact of firm size on corporate R&D investment, but it is still unclear whether this effect is linear. In this paper, we use the size of total assets to indicate the firm size.

4.2.4.2 Corporate age (Age). Younger companies and the companies that already have mature market experience have different judgments on market prospects, which will affect the company’s research engine. At the same time, the longer the company is established, the more likely it is to focus on BR investment. In this paper, we use the establishment time of each enterprise to indicate the age of the company.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Symbol</th>
<th>Ave</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total factor productivity</td>
<td>TFP</td>
<td>1.012</td>
<td>0.238</td>
<td>0.472</td>
<td>2.710</td>
</tr>
<tr>
<td>Basic research investment</td>
<td>BR</td>
<td>6.162</td>
<td>7.146</td>
<td>0.053</td>
<td>51.551</td>
</tr>
<tr>
<td>Human capital structure</td>
<td>HCS</td>
<td>0.562</td>
<td>0.224</td>
<td>0.105</td>
<td>1.000</td>
</tr>
<tr>
<td>Enterprise scale</td>
<td>Size</td>
<td>17.299</td>
<td>1.124</td>
<td>15.068</td>
<td>20.479</td>
</tr>
<tr>
<td>Enterprise age</td>
<td>Age</td>
<td>2.751</td>
<td>0.246</td>
<td>1.946</td>
<td>3.332</td>
</tr>
<tr>
<td>Capital structure</td>
<td>CS</td>
<td>0.415</td>
<td>0.209</td>
<td>0.022</td>
<td>0.964</td>
</tr>
</tbody>
</table>

**Note:** The variable of size and age is taken for the natural logarithm.
4.2.4.3 Capital structure (CS). When the total debt of the enterprise is far greater than the total assets, high interest squeezes the current operating cash flow of the enterprise, thus reducing the amount of funds invested by the enterprise in BR. At the same time, high interest rates on debt make companies face financial difficulties. If high-risk and long-term BR is blindly added to the project, the financial risks faced by companies will be further increased.

5. Results

5.1 Basic test

First, in order to avoid the pseudo regression, we perform a unit test before the regression analysis. According to the test results presented in Table II, all of the variables reject the null hypothesis that there is a unit root at the 1 percent significant level, which means that panel data are smooth.

Second, in order to avoid the result bias caused by multicollinearity, we test the correlation coefficient between variables. The test results are shown in Table III. It can be seen that all of the correlation coefficients between variables are less than 0.5, which means that there is no multicollinearity among the variables.

Finally, according to the Kao test and Pedroni test, we tested the long-term co-integration relationship between BR investment and corporate performance based on the PP and ADF statistic, and the test result is shown in Table IV. Observing the test result, we can confirm that there is a significant long-term balanced relationship between BR investment and corporate performance.

<table>
<thead>
<tr>
<th>Variable</th>
<th>LLC</th>
<th>Fisher–ADF</th>
<th>PP–Fisher</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFP</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.0000</td>
<td>Smooth</td>
</tr>
<tr>
<td>BR</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>Smooth</td>
</tr>
<tr>
<td>HCS</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td>Smooth</td>
</tr>
<tr>
<td>Size</td>
<td>0.0021</td>
<td>0.0011</td>
<td>0.0001</td>
<td>Smooth</td>
</tr>
<tr>
<td>Age</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td>Smooth</td>
</tr>
<tr>
<td>CS</td>
<td>0.0000</td>
<td>0.0003</td>
<td>0.0000</td>
<td>Smooth</td>
</tr>
</tbody>
</table>

Table II. Unit root test result

<table>
<thead>
<tr>
<th>Variable</th>
<th>TFP</th>
<th>BR</th>
<th>Age</th>
<th>Size</th>
<th>CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFP</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BR</td>
<td>0.0594***</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.1430***</td>
<td>-0.1478***</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>0.0143</td>
<td>-0.3732***</td>
<td>0.2841***</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>-0.0192</td>
<td>-0.4670***</td>
<td>0.1131**</td>
<td>0.3741***</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table III. Correlation coefficient matrix

Note: ***,***Significant at 5 and 1 percent levels, respectively

<table>
<thead>
<tr>
<th>Method</th>
<th>Statistics</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kao test</td>
<td>ADF</td>
<td>-6.850***</td>
</tr>
<tr>
<td>Pedroni test</td>
<td>Panel PP statistic</td>
<td>-17.963***</td>
</tr>
<tr>
<td></td>
<td>Panel ADF statistic</td>
<td>-13.164***</td>
</tr>
<tr>
<td></td>
<td>Group PP statistic</td>
<td>-16.579***</td>
</tr>
<tr>
<td></td>
<td>Group ADF statistic</td>
<td>-11.379***</td>
</tr>
</tbody>
</table>

Table IV. Cointegration test result

Note: ***Significant at 1 percent level
5.2 The relationship between basic research investment and corporate performance

Based on the sample data, we plot the scatter plot between BR investment and corporate performance, as show in Figure 1. Observing the fit curve in the chart, we can see that there is a positive one-way function relationship between BR investment and corporate performance. From this, it can be seen that for the current stage, the increase of BR investment in enterprises is an effective measure to enhance corporate performance.

On this basis, based on the previous formulae (1) and (2), this paper quantitatively analyzes the correlation between BR investment and corporate performance. Before conducting empirical analysis, the specific form of the regression model needs to be determined. The role of the individual effect $\mu_i$ in the panel model is to control the enterprise’s fixed effect that does not change with time. If the unobservable variable $\mu_i$ is not related to the explanatory variable, including it into the error term does not affect the consistency of the estimation result, thus the regression model can be estimated by the random effect method. Otherwise, the fixed effect approach is adopted. Hausman test result shows that the $\chi^2$ statistic is 12.20 and rejects the null hypothesis of random effect at a 5 percent significance level, so a fixed effect method is used to eliminate the effect of $\mu_i$. In addition, in order to test the robustness of the influence of BR investment on corporate performance, we used a stepwise regression method and added control variables in order. The final regression results are shown in Table V.

From the first column of Table V, it can be seen that the BR has a significant positive effect on corporate performance, with an elasticity coefficient of 0.118, which means that BR can generally promote the improvement of corporate performance. In the Columns 2–4 of Table V, this paper gradually adds the control variables such as enterprise age, enterprise size and the CS, and the influence of BR on corporate performance is tested again. From the estimation results of each column, it can be seen that BR still has a significant role in promoting the corporate performance. This conclusion confirms the correctness of the first hypothesis in this paper. That is, the investment in BR has a positive effect on the improvement of corporate performance. Consistent with the past research, previous authors have pointed out that innovation is an important way for late-growth enterprises to leapfrog development, and BR can accumulate corresponding innovation capabilities for enterprises, thus improving corporate performance (Martinez-Senra et al., 2015). Meanwhile, this paper believes that BR can improve corporate performance through enhancing the absorptive capacity and

![Figure 1. Curve fitting](image-url)
increasing the synergy of the company (Griliches, 1986). It needs to be noticed, China’s BR investment, especially the BR of enterprises, is still lower than the general level of emerging countries and developing countries at present, and also lower than the level of industrialization in the same period in developed countries. The lag of financial development, especially the slow development of capital markets, the lack of R&D institutions and researchers, and the lack of intellectual property protection all contribute to the low proportion and strength of BR in China. Therefore, improving the support and allocation efficiency of government financial support for BR gives full play to the complementary role of government plans and market mechanisms, and induce enterprises to increase investment in BR (Gaur et al., 2018). Other factors include coordinating all aspects of the company’s own factors, integrating global knowledge to improve BR capabilities, and achieving the learning and absorption of cutting-edge knowledge and technology (Lee et al., 2017; Popli et al., 2017).

In order to fully consider the time lag effect of corporate performance, this paper establishes a dynamic panel measurement model and introduces the lag item of the corporate performance as an independent variable into the model. On this basis, the impact of potential factors not included in the econometric model on corporate performance is examined, and then we compared with the regression results obtained based on the static panel model. For dynamic panel estimation, this paper mainly uses the generalized method of moments (GMM) estimator proposed by Arellano and Bover (1995) and Blundell and Bond (1998), and the specific results are shown in Columns 5 and 6 of Table V. According to AR test and Sargan test results listed in the last three rows of Table V, the serial correlation test of residuals AR(2) proves that the null hypothesis is accepted at the 10 percent significance level. Besides, the values of Sargan test are greater than 0.1, implying that the instrument variables are available. According to the regression coefficients of the lagged corporate performance (L.TFP), the regression coefficients are all in the interval [0, 1] and pass the test of significance at 10 percent level, and positive regression coefficients indicate that there is an obvious lag effect of corporate performance. It means that the growth of corporate performance has strong time inertia, that is, the changes in early-stage corporate performance has a significant impact on current growth (Bischoff and Buchwald, 2018). According to the regression coefficients, the impact of BR investment on corporate performance is still significant positive, and the elasticity coefficients under system GMM and differential GMM are 0.066 and 0.266, respectively.

### Table V.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Static</th>
<th>Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>L.TFP</td>
<td>0.069*** (3.96)</td>
<td>0.058*** (3.19)</td>
</tr>
<tr>
<td>BR</td>
<td>0.633*** (2.32)</td>
<td>1.030*** (2.52)</td>
</tr>
<tr>
<td>Age</td>
<td>−1.020 (−1.30)</td>
<td>−0.939 (−1.19)</td>
</tr>
<tr>
<td>Size</td>
<td>−0.039 (−1.03)</td>
<td>0.007 (0.10)</td>
</tr>
<tr>
<td>CS</td>
<td>−0.100*** (−4.17)</td>
<td>1.795 (0.92)</td>
</tr>
<tr>
<td>cons</td>
<td>0.045</td>
<td>0.061</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.583</td>
<td>0.316</td>
</tr>
<tr>
<td>AR(2)</td>
<td>0.188</td>
<td>0.07</td>
</tr>
<tr>
<td>Sargan test</td>
<td>0.316</td>
<td>0.287</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** The numbers presented in the parentheses are the $t$-statistic of the regression coefficients. *, **, *** Significant at 10, 5, and 1 percent levels, respectively.
5.3 The moderating effect of human capital structure on the relationship between basic research investment and corporate performance

In the previous section, we have confirmed the positive impact of BR investment and corporate performance. At the same time, we pointed out that the relationship between R&D and corporate performance may be quite different due to the influence of the company’s own factors (Lee et al., 2017; Popli et al., 2017). Combined with the theoretical analysis of the previous article, this section analyzed the moderating effect of HCS on the relationship between BR and corporate performance, this paper adds the interaction term of HCS and BR into the basic model, as shown in Equation (5). It is noteworthy that this model incorporates the interaction term of HCS, which may lead to multicolinearity. Thus, this paper uses the decentralization method to process the relevant variables (Zhu et al., 2017). Similarly, in the empirical analysis process of interaction effect, we test from the static and dynamic two perspectives, in which the static model also adopts a fixed effect model and gradually adds control variables. In the dynamic model, we still show the regression results of the system GMM and differential GMM models. The specific results are listed in Table VI.

According to the estimation results in Table VI, it can be seen that after the interaction items are added, the effect of BR on corporate performance is still significantly positive. The regression coefficient of the interaction term is significantly positive, which means that in the companies with a sound HCS, the positive effect of BR on corporate performance is more significant, as the effect of BR on corporate performance of low HCS and high HCS is 0.056 and 0.107, respectively. When we successively add the control variables into the regression model, the regression coefficients of the interaction item are always significant. In addition, in the dynamic panel model which added the first-order lag items of corporate performance into the regression model, the regression results of system GMM and differential GMM confirm that the positive regulation effect of HCS is robust. To sum up, the second hypothesis in this paper is proved, that is, the HCS positively regulates the relationship between BR investment and corporate performance. Hence, we believe that better human capital not only has a positive influence on corporate performance (Hazen et al., 2017), but also is an important dynamic capability embedded in corporate organizations. That is, to a certain extent, BR investment itself is static and cannot be automatically converted into an enterprise value. Economic performance and competitive advantage must be achieved through the management and creation of knowledge.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Static (1)</th>
<th>Dynamic (2)</th>
<th>Differential (3)</th>
<th>(4) System GMM</th>
<th>(5) Differential GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.TFP</td>
<td>0.056*** (2.98)</td>
<td>0.049*** (2.58)</td>
<td>0.048*** (2.58)</td>
<td>0.047** (2.46)</td>
<td>0.052* (2.17)</td>
</tr>
<tr>
<td>BR</td>
<td>0.051** (2.13)</td>
<td>0.040* (1.63)</td>
<td>0.038* (1.59)</td>
<td>0.036* (1.52)</td>
<td>0.052* (1.74)</td>
</tr>
<tr>
<td>BR*HCS</td>
<td>0.525* (1.88)</td>
<td>0.890** (2.13)</td>
<td>0.964** (2.26)</td>
<td>0.373** (2.21)</td>
<td>0.302 (−0.28)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>−0.921 (−1.18)</td>
<td>−0.859 (−1.09)</td>
<td>−0.410 (−0.69)</td>
<td>−1.112 (−0.50)</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td>−0.033 (−0.87)</td>
<td>−0.028 (0.80)</td>
<td>0.105 (0.87)</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>−0.125*** (−4.70)</td>
<td>−0.631** (−2.33)</td>
<td>1.632 (0.84)</td>
<td>1.348 (0.68)</td>
<td>0.724 (0.44)</td>
</tr>
<tr>
<td>cons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.059</td>
<td>0.068</td>
<td>0.072</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.002</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(2)</td>
<td>0.380</td>
<td>0.612</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sargan test</td>
<td>0.247</td>
<td>0.639</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table VI. Moderating effect regression results

Notes: The numbers presented in the parentheses are the t-statistic of the regression coefficients. ***,***Significant at 10, 5, and 1 percent levels, respectively
In particular, under the condition that the company’s BR investment is relatively general, the improvement of the human resource capital can make up for the lack of insufficient resources (Salvato and Vassolo, 2018). Especially in the digital age, effectively adjusting the structure of human capital and improving the ability of acquisition and processing of information (Tseng, 2017) has a significant positive promote effect on BR investment and finally promoting the improvement of corporate performance.

6. Conclusion and implication

With the wide implementation of technology introduction and encouragement of innovation policies, some Chinese enterprises have achieved a transition from technology introduction to independent innovation. However, from the current stage, the enterprise innovation in China mainly based on improved innovation, the number of invention patents is far lower than the number of design patents and utility model patents, and the independent innovation capability related to core technology is still weak. Since the beginning of the twenty-first century, BR has become an important channel for expanding the breadth and depth of applied knowledge, and an important driving force for promoting Chinese enterprises to achieve independent innovation and enhance their core competitiveness. If China wants to successfully complete the dynamic transition from technologically catch-up development to self-innovation development and avoid falling into a low-level cycle of technological progress in “introduction–improvement–reintroduction,” it is urgent to increase investment in BR, especially with high technology. Investment in BR related to enterprise development will increase the self-innovation capabilities and reduce the dependence on core technology imports so as to achieve a high level of “introduction–improvement–independent development,” and then improve the corporate performance and core competitiveness.

In this paper, the authors elaborate theoretically the influence mechanism of BR investment on corporate performance at first. At the same time, the authors discuss the moderating effect of HCS based on dynamic capability theory. At the empirical research level, this paper expands on the basis of the Mansfild model and takes the form of a double log-linear model to analyze the influence of BR investment on corporate performance. On this basis, the authors collect a large-scale data from Chinese enterprises listed in A-share and test the rationality of the theoretical hypothesis from static and dynamic panel model two perspectives. The results show that: BR shows a significant positive effect on corporate performance. It means that BR serves as an important channel for knowledge expansion, and the promotion of corporate performance shows the effect of knowledge accumulation. HCS moderate the positive impact of external collaboration on environmental and economic performance. It means that in the companies with a sound HCS, the positive effect of BR on corporate performance is more significant. This paper has profound policy implications:

1. BR has significantly promoted enterprise performance, and in most cases, the output elasticity of BR exceeds the output elasticity of R&D personnel. BR enables enterprises to break through the technical problems, and effectively help companies to change their original technology routes, discover and grasp new technological opportunities, and establish “new production functions” to significantly improve performance. For now, there is still much room for growth in the absolute amount of BR investment for Chinese companies, and the potential for BR to improve corporate performance remains to be further released. Therefore, the state encourages the policy orientation of BR, and should also face the innovation of enterprises, and encourage enterprises to play a greater role in the main body of innovation. Enterprises need to start with BR project investment, provide more equipment and financial support for researchers engaged in BR, and increase the research enthusiasm of researchers to gain more BR results.
Human capital is an important factor that can bring competitive advantages to enterprises. Enterprises must improve the dynamic capabilities of knowledge acquisition, integration, and release of human capital. As far as the improvement of corporate performance is concerned, both the upgrading of human capital of enterprises and the improvement of the management and conversion capabilities of basic research investment can be achieved. In particular, managers should reasonably construct and optimize the enterprise team, select the people with high education and rich functional background to join the team, which will help to obtain more BR investment. Other factors include building a learning organization, providing training and continuing education opportunities, enabling an effective information exchange platform within the team, strengthening the team’s cohesiveness, and enhancing the stock and quality of the team’s human capital; strengthening corporate culture, stimulating employees’ sense of mission and belonging, and forming a good organizational atmosphere within the enterprise.

In the digital age, based on the support of modern multimedia and information technology, it is necessary to build a digitization management platform within the enterprise, employ digital technology to realize the digital management of human capital, and quickly gather information and provide timely and effective analysis. In addition, the full coverage of digitization management should be achieved, such as talent recruitment, promotion, etc.; thus enhancing core competence and improving the performance of the enterprise.

Acknowledgments
The authors thank anonymous referees and editors for their very constructive comments on the initial draft of the paper, and the work was support by the National Natural Science Foundation Project (Grant No. 71303029), Major Projects in Philosophy and Social Science Research from the Ministry of Education of China (Grant No. 14JZD031), Dalian Science and Technology Plan Project (2015D12ZC176), and Dalian Youth Science and Technology Star Breeding Plan Project (2016RQ004). However, the authors are responsible for any errors.

References


Further reading


Corresponding author

Xiongfeng Pan can be contacted at: xiongfengpan@163.com

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Role of cloud ERP and big data on firm performance: a dynamic capability view theory perspective

Shivam Gupta
Montpellier Business School, Montpellier Research in Management, Montpellier, France
Xiaoyan Qian
Department of Computer Science and Engineering, Southern University of Science and Technology of China, Shenzhen, China
Bharat Bhushan
School of Management, XLRI – Xavier School of Management, Jamshedpur, India, and
Zongwei Luo
Department of Computer Science and Engineering, Southern University of Science and Technology of China, Shenzhen, China

Abstract
Purpose – Technological developments have made it possible for organizations to use enterprise resource planning (ERP) services without indulging in heavy investments like IT infrastructure, trained manpower for implementation and maintenance and updating the systems regularly to maintain business competitiveness. Plug and play model offered by cloud ERP has led to a constant creation of large data sets which are structured, semi-structured and unstructured by nature. Thus, there has been a need to analyze such complex data sets and the purpose of this paper is to focus on how cloud ERP and big data predictive analytics (BDPA) will impact the performance of a firm.

Design/methodology/approach – A dynamic capability view (DCV) theory-based model was developed and the authors have collected data by using an online questionnaire from India. Thereafter, the authors have analyzed it by employing structural equation modeling.

Findings – SEM analysis of 231 respondents showcases that the use of DCV theory to define the relationships of cloud ERP and BDPA has been the right move. Out of the 13 hypotheses empirically tested, only 7 hypotheses were supported by the data.

Research limitations/implications – The study showcases cross-sectional data from India. It would be interesting for this study to see if the country-level differences would influence these relationships between cloud ERP and financial performance, BDPA and financial performance and cloud ERP and BDPA.

Originality/value – This study empirically tests the relationship of cloud ERP and BDPA through a model based on DCV theory.

Keywords Big data, Organizational culture, Structural equation modelling, Firm performance, Cloud ERP, Dynamic capability view

Paper type Research paper

1. Introduction
Enterprise resource planning (ERP) systems provide firms extensive facilities and capabilities to share and transfer data and processes of organizations inside and outside the enterprise into a single system and single database (Peng and Nunes, 2013). Sharing data...
between firm’s departments or firms across the supply chain helps in many aspects \((\textit{inter alia} \text{ in the form of decision support})\) and aims to achieve the objectives of better firm performance (FP). Elmonem \textit{et al.} (2016) commented ERP as a category of business management software system that aims to integrate all functional units, typically a suite of integrated applications in a cooperative way. It facilitates organizations to collect, record, manage and interpret data from these business activities. The fact is that, ERP has so far been widely implemented by different organizations with different sizes in many sectors and in many countries to seek competitive advantages in the market.

With time going by, ERP has revolutionized, and continuous upgradation has taken place to strengthen its functionality of resources sharing and integration capabilities of functional units. Scholars increasingly tout internet-enabled ERP systems as an important perspective for the performance of a firm or even firms across the supply chain. With the advent of cloud computing technologies in the late 2000s, Peng and Gala (2014) highlight that there exists an increasing trend for firms to move their ERP-based applications and database into the cloud. According to Salleh \textit{et al.} (2012), cloud ERP as a concept has been a boon to the FP, \(\textit{inter alia}\) for small and medium enterprises over large companies since they could conform to the infrastructure requirements of the on-premise ERP solution as well as the high cost. Cloud computing brings firms the very model that enables ubiquitous access to share data and resources to achieve coherence, get the application up and run faster, often over the internet.

Exploring the cloud enterprise resource planning (CERP) system enabled with predictivity ability may help to resolve high uncertainties and gain more competitive advantages than other competitors in the dynamically changing market. According to Duan \textit{et al.} (2013), CERP systems give the enterprise a chance to access the advanced computing resources that are available over the cloud, and even support the firms to manage their business functions to achieve higher productivity. Beheshti (2006) also argued that CERP systems are capable to manage and handle the large volume of operations and information that is created daily within the firm. Besides the potential benefits for operational performance (OP), one of the main drivers from a CERP would be the technical and operational integrations of functional processes to harmonize the data and information stream based on product lifecycle (PLC) (material flow of goods or services) (Qian \textit{et al.}, 2016). In view of Beatty and Williams (2006), this would happen through integrating the values across PLC within a seamless business process streamlining, which could potentially precede the firm’s market competitiveness and responsiveness in the rapidly changing environment. Therefore, implementing ERP systems on the cloud platform is showcased for resolving the limitations of ERP systems and provides better scalability, reliability, availability and cost efficiency that are all the very components for a higher FP.

On the other hand, big data predictive analytics (BDPA) would be the next big frontier of innovation, efficiency, productivity and competitiveness of an organization (Srivastava, 2014; Waller and Fawcett, 2013). Reaping the benefits of big data, CERP systems enhanced with e-commerce capabilities and its ability of integration and sharing resources and capabilities, collaboration with corporate alliance (suppliers, partners and even customer portals), and tracking of incoming resources and outgoing final products extends the visibility and control from inside and outside with big data analytics (McAfee \textit{et al.}, 2012). In doing so, upstream and downstream firms in the supply chain could provide reporting capabilities to management via sharing information (i.e. data) needed to support strategic decision-making that is of huge importance for long-term benefits of firms in the supply chain. Organizations in various sizes from various areas are jumping on the bandwagon of big data and predictive analytics (BDPA) due to the data sharing and interactive nature of CERP systems for firm competitive advantages.

By definition, BDPA is a decision-making field which consists of big data; use various statistical tools and techniques and machine learning, deep learning artificial intelligence and
even data mining to derive potential insights from huge data sets, improving the market performance (MP) and OP of a firm (Gupta et al., 2018a). Analyzing big data using predictive techniques would be a necessity for decision support, although big data alone is ubiquitous (Prescott, 2014; Duan and Xiong, 2015). Extracting potential insights from large data sets via analytics techniques are an all-encompassing term among the senior executives to make decisions for their enterprises. McAfee et al. (2012) highlight that firms of various sizes need to take the data-driven decision-making into strategy practices via which the top management decision-makers execute any plans based on data instead of gut feeling. Apart from data, Waller and Fawcett (2013) argue that the appropriate managerial skills (MS) and technical skills (TS) play a non-substitutable role in the success of predictive analytics initiatives. Matthias et al. (2017) include that the application and exploitation of BDPA would create first-mover competitive benefits for sustainable improvement for a firm. Consistent with the rich research around (surrounding) the extraction of valuable data and potential insights from a large database, we argue that BDPA may be a well-accepted technology for decision-making on the performance of a firm. For instance, organizations may react differently to the same CERP systems on the FP due to the differences in the ability of BDPA capability.

The existing research and studies on CERP and BDPA, respectively are very rich, but the vast majority of them focus merely on its own capabilities and effects. In contrast, knowledge on the joint role of CERP and BDPA capabilities for the performance of a firm is scant. Very few scholars investigated on the specific relationships of CERP and BDPA, CERP and FP, BDPA and FP. Such a void leaves a significant gap between firm’s resources and capabilities and its performance. To address this gap, we propose a theory-driven and empirically proven model that aligns CERP and BDPA and could explain the impacts of CERP and BDPA on FP drawing on dynamic capability view (DCV) model (Teece et al., 1997). More specifically, our current research objectives of this paper are to address two main issues as follows:

1. develop a theoretical framework based on DCV theory to understand the role of cloud-based ERP services and BDPA on the performance of a firm; and
2. empirically validate this theoretical framework by employing structural equation modeling (SEM).

This paper is organized as follows. In Section 2, we begin with a brief review of the relevant literature pertaining to CERP and BDPA. In Section 3, we describe our theoretical model based on DCV and hypotheses development. In Section 4, we would outline the research methodology and data analytics for the empirical validation of this theoretical framework by employing SEM. Section 5 consists of our discussion related to our analysis results, including theoretical contribution, managerial implications and limitations and further directions. In Section 6, we conclude with our discussion results.

2. Theoretical background

2.1 Dynamic capability view

In the strategic management area, the distinct mechanism of capability-building resource-based view (RBV) is for understanding how we could create our economic competitiveness for a firm. According to Teece et al. (1997), in order to locate a firm’s position in the market via its resources, the RBV theory was the first concept among the management strategy literature. Wernerfelt (1984) supplemented the conception that resources could be studied and examined as the source of a firm with competitive advantages. RBV tends to support firms to create more economic advantages than their competitors by being more effective at defining internal and outside resources and deploying strategic resources, and subsequently building capabilities (Makadok, 2001; Kim et al., 2015). However, Kraaijenbrink et al. (2010) asserted that RBV was not able to...
address how firms utilize resources and capabilities in a dynamic market as RBV is primarily a static theory that helps the firms to maintain a competitive advantage by employing resources at their disposal. Therefore, the DCV as an extension of RBV (Wang et al., 2016) was proposed which promotes innovation (Lawson and Samson, 2001).

Going back to the origins of capabilities, it began with the request that static nature of RBV could not fully showcase how the resources of a firm developed and integrated in a rapidly changing market (Teece et al., 1997; Winter et al., 2003; Smith et al., 2014). Teece et al. (1997) defined the dynamic capabilities as the abilities to deploy, integrate, build and reconfigure the competencies inside and outside a firm to resolve the dynamically changing market. Lawson and Samson (2001) highlighted dynamic capabilities support enterprises to improve the profits by managing firm’s capabilities (efficiency, quality, velocity, flexibility, etc.) in a dynamic and uncertain environment. Given the need of rapid responsiveness of a firm’s resources stock to increasingly dynamically changing environments, Vogel and Güttel (2013) indicated that dynamic capabilities would be of inherent strategic relevance to a firm, to keep pace with competitive dynamics. As such, DCV theory would be regarded as the distinct process that allows resources and focuses on learning and change capabilities to relate them to FP.

Given that one of the main objectives of this study is to identify various resources that will enable firms to create CERP and BPDA capabilities, which in turn may lead to superior MP and OP, the choice of DCV as a theoretical framework for this study seems appropriate.

In this paper, DCV is used to conceptualize BDPA and CERP as capabilities that have an impact on FP. Resources like data (D), MS) and TS support BDPA (capability), which impacts on MP and OP. In a similar manner, dynamic capabilities grouped into organizational factors (OF), people factors (PF) and technological factors (TF) constitute CERP (capability) that has an impact on MP and OP of an organization (Figure 1).
2.2 Cloud-based ERP (CERP)
Most recently, one of the most popular trends is cloud computing that has a huge potential to reshape the way ERP systems operate. Armbrust et al. (2010) defined cloud computing as both the applications and systems software in the information centers. Cloud computing would be a key computing paradigm for the next ten years, anticipated by Smith et al. (2014). Literature has highlighted the role of ERP in the cloud computing enables many applications of web services via the internet (Chen et al., 2015).

It is not surprising that toward developing a fully functional CERP system has been a center of focus by both firms and researchers, given that CERP is built to provide firms with huge benefits of flexibility, improved accessibility, scalability, lower upfront and operating costs, rapid implementation, cost transparency, sales automation, higher security standards and free trials (Gupta and Misra, 2016). The power of CERP usually refers to the performance of an organization, including MP and OP. These benefits derived from CERP system not only saves the cost of operation when improving the effective productivity but also supports the changed business size into the dynamically changing market when satisfying the firm’s needs of new markets shares more quickly than competitors.

2.3 Toward the conceptualization of CERP capability
The key competitive edge for a firm now is its capability to define, standardize and adapt its processes and information with supplier, partner, customers based on PLC in a dynamically changing environment (Chen et al., 2015). Cloud computing allows firms convenient and on-demand access to share a pool of configurable resources. Therefore, cloud computing supports to improve the operational efficiency and help firms to achieve dynamic capabilities (Battleson et al., 2016).

In this paper, we argue that CERP could be conceptualized as a capability with the aid of cloud computing. This paper would identify the dynamic capabilities that explain how firms could effectively respond to market dynamism by developing CERP capability. CERP capability is created by the intrinsic factors combination of OF, PF and TF (Gupta and Misra, 2016) for they are under the control of cloud user. In terms of the determinants of CERP capability, OF, PF and TF are briefly discussed below.

2.3.1 Organizational factors (OF).
The research proposed that the success of CERP capability building was affected by OF (Law and Ngai, 2007) and suggests that OF act as the major determinants of FP (Hansen and Wernerfelt, 1989). Determining the appropriate organizational factor or the construct of performance involves communicating the top management strategy to front-line subordinates and organizational structures and systems to people simultaneously. The OF are considered here to capture the multi-dimensional phenomena – strategic goals and objective, implementation strategy, business process re-engineering, organizational resistance, project management and budget and even communication.

The important organizational factor in CERP is the alignment between CERP and organizational objectives. The fit between strategic goals and objectives and CERP systems is brutal critical to achieving FP (Law and Ngai, 2007). In terms of the implementation strategy, it would be set before the implementation of CERP for it guides the future function and capability for our CERP capability. The project budget has a positive impact on better outcome for the enterprise as it is in sync with the implementation strategy (Hasibuan and Dantes, 2012; Somers and Nelson, 2004). Business process re-engineering is the extent to which organizations have to change and re-engineer the existing business process to fit the coming new system in line with the requirements of customization or new markets development (Saleh et al., 2013). What is asserted by prior research is that more the abilities an organization holds for business process re-engineering to change, more powerful could
be its ERP systems (Grover et al., 1995). Consequent with re-engineering business process in CERP solutions, organizational resistance may be also be a hindrance factor to operate smoothly (Utzig et al., 2013). Project management is indispensable role for the CERP capability, which refers to set the vision and directions for business process and harness the cooperation and potentials of employees to exploit the technological capabilities of CERP (Al-Mashari et al., 2003; Esteves and Pastor-Collado, 2001), as well as efficient execution of implementation strategy (Gupta and Misra, 2016). Communication has to be in sync with the project management for better understanding the roles between superior–subordinate and among employees with good working relationships across the project. Most projects failed due to its communication, either a lack of thereof or miscommunication. The need for effective communication is permanent and would affect all the factors listed above. All these measurement constructs would be in the collaboration as OF in executing CERP system.

2.3.2 People factors. People are the biggest potential power of a firm. According to Lalsing et al. (2012), PF have been proven to be the most critical success factors for development systems. From the employee’s perspective, there would be many measure constructs identified to cover the PF. The key components which are being created as PF are user involvement, vendor selection, project team, top management support, training of user and trust on vendor (Gupta and Misra, 2016).

During building CERP capability, there would be a non-substitutable consideration that actively involves user (Françoise et al., 2009; Hasibuan and Dantes, 2012). User involvement and participation could help to ensure user’s requirements with our better quality and user-friendly CERP systems since user involvement in the decision-making process result in greater attachment to the CERP capability and functionality (Lalsing et al., 2012; Françoise et al., 2009). In addition, training of user is also perceived. Needless to say, our CERP system would not create any values and profits if the employees do not know how to operate it. The need for user training and support is crucial in building CERP capability.

Vendor selection is an important decision regarding building CERP capability since most organizations purchase ERP packages from vendors. Based on the end-users requirements and owners’ consultants, organizations could take guidance to map the needs of the organizations with the respective vendor (Saleh et al., 2013; Françoise et al., 2009). In addition, trust on the vendor is a necessity to keep the good working relationship between firms and cloud vendor. Vendor support provides extended technical assistance for emergency aid and maintenance, updates, service responsiveness and user training (Somers and Nelson, 2001) during the period of CERP existence.

Project teamwork refers to the amount of knowledge and skills that are responsible for CERP capability building and business operation process. Umble and Umble (2002) suggested ERP teams should be composed of cross-functional members who possess skills, good reputations on past accomplishments and decision-making responsibility. In addition, the team would be supported by the leadership of the company.

Literature underlines the role of top management in orchestrating resources and creating capabilities and subsequently helping to achieve the competitive advantages of a firm (Hitt et al., 2016; Chadwick et al., 2015). Top management not only defines new objectives that could provide employees a clear vision of the orientation the organization is taking with careful consideration to the objectives but also supports all the decisions that need to be made to handle any conflicts that may happen. Thus, top management is required to commit and support for CERP system with enthusiasm, full consideration and even continuously monitoring among all the process (Prajogo and Olhager, 2012).

2.3.3 Technological factors (TF). Factors related to technologies that shall be affecting building CERP capability would be considered for the selection of ERP packages, IT infrastructure, data integrity and system testing, and its functionality of
cloud ERP modules. Selecting ERP packages based on different cloud layers from a vendor should be strategic in nature in such a way that it matches the required business process (Gupta and Misra, 2016) and enhances the organization’s competitiveness and efficiency. IT infrastructure refers to the important components required for the CERP’s existence, operation and management in the form of hardware and software (Alaskari et al., 2012; Somers and Nelson, 2001), which is more significant to CERP vendor rather than the user. Data integrity is a critical aspect to the design, implementation and usage of CERP systems that are required to store, process and even retrieve data, as well as for the maintenance and the assurance of the accuracy and consistency of data over its entire lifecycle (Boritz, 2005). To make sure that CERP system would operate quickly and smoothly after the go-live, it is often tenable that system testing would continue as long as errors remain. Functionality is supposed to be in the sync with the selection of ERP vendor since it makes sure the consistency between ERP modules and organization business requirements. The decision for all these TF needs to be done before building the implementation of CERP (Gupta and Misra, 2016).

2.4 Big data and predictive analytics
Given business processes with CERP capability and data sharing with the up and down partners are moving online, large-scale data would be created from these applications. As what most researchers asserted, BDPA would be the next big thing for firms to gain competitiveness in the dynamically changing market (Akter et al., 2016; Wamba et al., 2015). BDPA is actually an interdisciplinary field due to leveraging not only the statistical technologies such as regression, time-series analysis, etc., but also the computer and data science tools including data mining, machine learning, etc. (Dubey and Gunasekaran, 2015). It would be defined as a systematic process of descriptive analytics for explaining the data rules, predictive analytics for picture future insights and the final prescriptive analytics for optimizing or simulating the outcomes of organizational decisions.

In addition, recent scholars have acknowledged that BDPA is an organizational capability that they would process and exploit to know how organizations could achieve and sustain competitiveness regarding MP and OP of a firm (Gupta and George, 2016; Wamba et al., 2017). So how organizations could exploit resources and capabilities to build a BDPA capability would be examined next, which is defined as a firm’s edge to assemble, integrate and deploy its big data-specific resources to gain market share or improve profitability. Drawing upon DCV logic, we suggest that firms need a unique combination of data, MS and TS as the resources of building a firm-specific BDPA capability for making operational decisions or predictions.

2.5 Toward the conceptualization of BDPA capability
2.5.1 Data (D). The world is witnessing an unprecedented huge interest in big data that is heralded as the next big hit for firms to gain the competitive edge (Frisk and Bannister, 2017; Rajput and Singh, 2018). The term “big data” is often used to describe a resource that features big in volume, big in forms (structured data, unstructured data and often semi-structured data) and big in velocity (fast-changing and real-time streaming), which most firms could approach (Lamba et al., 2018). Gupta and George (2016) highlighted that data are also the premise of deriving usable information for improved decision-making, action and positive change, besides labor and capital. Nevertheless, data by itself do little value to organizations. In other words, big data on its own are unlikely to be a source of competitive edge, since most firms have likely collected hordes of structured, unstructured or semi-structured data from various sources (Lamba and Singh, 2018a). It is imperative to have sophisticated data administration, data analytics and processing techniques to extract inherited insights (Beyer and Laney, 2012). Data are one of such immense resources, which are necessary but not
sufficient to create a BPDA capability. It is imperative for firms to be aware of the various resources that are required to build BPDA capability (Lamba and Singh, 2018b).

2.5.2 Managerial skills. MS are developed as a result of long years working experiences, which play a non-substitutable role for analytics projects as managers. The success of BDPA projects greatly depends on how well managers could infuse employees the common goals and assemble a team with right skills (Lamba and Singh, 2017; Dubey et al., 2018). The essential quality to predict market behavior and the interpersonal skills to develop swift have been regarded as the critical parts to the successful use of BDPA for FP. Big data analytics managers should be enabled to work with functional managers, suppliers and customers, to coordinate big data-related activities, to anticipate the future business needs with the good sense of where to apply big data and to understand and evaluate the output extracted from big data (Gupta and George, 2016).

2.5.3 Technical skills. TS commonly refers to the know-how to possess specific skills and ability to extract intelligence from big data with the knowledge in statistics, computer and data science, as well as problem-solving skills and strong people skills (Lamba and Singh, 2016; Jeble et al., 2018). In terms of developing TS, firms could hire new talented employees with BDPA capability or conduct some big data analytics training for current employees. Big data analysts need are supposed to have the rights skills to accomplish their jobs smoothly with the suitable education and work experience (Schoenherr and Speier-Pero, 2015). More specifically, these right skills involve competencies and proficiency in statistics analytics, data cleaning, extraction analytics, data mining, machine learning and master of programming paradigms (Davenport, 2014).

2.6 Organizational culture

Previous researchers have acknowledged the intangible resource of organizational culture is a source of sustained FP since it would be built over a long period and varies from organization to organization, which could not be duplicated by other competitors or coordinators (Teece, 2015; Jeble et al., 2018). Along similar lines, recent work regarding big data has identified organizational culture as a critical success factor to inhibit an organization’s ability to benefit from big data for analytics and predictive projects (Ross et al., 2013), urging firms to develop data-driven culture (LaValle et al., 2011; Ross et al., 2013; Gupta and George, 2016). To fully realize the potential of big data origins from firms, it is critical and necessary to develop a data-driven organizational culture.

Organizational culture can be viewed as “corporate personality” as it encompasses collective values, beliefs, behaviors and principles of organizational members that contribute to an independent enterprise with unique environment (Needle, 2010). Though organizational knowledge would never wear out, but it tends to become outdated. In line with the data-driven culture, organizational members (including top decision-maker and executives, middle managers and lower-level employees) ought to concert efforts to exploit their existing knowledge and explore new knowledge that would upgrade them. Their effort needs to be in place to keep learning the latest knowledge from internal and external environment. It is essential because the competitive environment keeps changing in the economics, technology, management, politics and even society. Based on the works of Gupta and George (2016) and Ross et al. (2013), organizational culture is the key intangible resource to contribute to building BDPA capabilities.

Prior literatures have proposed several ways to classify organizational culture into either relation-oriented or transaction-oriented culture (McAfee et al., 2002) or control and flexible orientation regarding organizational culture (Khazanchi et al., 2007). Control orientation values the predictability and efficiency, leveraging core competencies, profile valuation, value–practice interactions and value congruence (Liu et al., 2010; Dubey, Gunasekaran, Childe, Papadopoulos,
Luo, Wamba and Roubaud, 2017). In contrast, flexible orientation emphasizes innovation paradoxes, creativity, spontaneity and risk-taking (Liu et al., 2010; Dubey, Gunasekaran, Childe, Papadopoulos, Hazen, Giannakis and Roubaud, 2017; Dubey, Gunasekaran, Childe, Papadopoulos, Luo, Wamba and Roubaud, 2017). Following with the works of Dubey, Gunasekaran, Childe, Papadopoulos, Hazen, Giannakis and Roubaud (2017) and Liu et al. (2010), our study adopted the paradoxical orientations of control and flexibility for the moderating effects of organizational culture study.

2.7 Firm performance
To fully measure the difference in the FP among organizations, we have considered two distinct factors, namely, MP and OP. Based on prior academic literature, FP comprises these specific actual outcomes of a firm: MP (market share, etc.), OP (profits, return on assets, etc.), shareholder return, customer service, social responsibility and even employee stewardship (Richard et al., 2009; Upadhaya et al., 2014). MP and OP are the two important distinct dimensions or components standard to account for organizational effectiveness or to measure the performance of a firm (Gupta and George, 2016; Rai and Tang, 2010; Wu et al., 2015; Wang et al., 2012). Consistent with the previous literature on FP, we select these two separate dimensions of FP in our study (i.e. MP and OP).

2.7.1 Market performance. In this study, we consider MP as an important role that acts on FP. The control variables to MP will be embodied in the abilities to explore market more quickly, to introduce new products or services into the market faster, and to gain higher success rate of new products or service and more market share than other competitors (Ji-fan Ren et al., 2017).

2.7.2 Operational performance. In our study, we consider OP as another important part. The control variables of OP will be presented on the benchmark of productivity, profit rate, return on investment (ROI) and sales revenue. An organization with higher OP would mean more effective productivity, higher profit rate and ROI and more sales revenue than that of other competitor.

3. Theoretical framework and hypotheses development
In order to develop a theoretical framework to evaluate how BDPA and CERP act on FP, this study began by investing commonly cited resources that influence CERP and BDPA capabilities perception following DCV logic that explains an organization’s competencies and competitiveness in dynamically changing environments. Resources and capabilities are the core components of DCV: resources refer to technology, people and organization whereas capabilities represent a special type of resource that aims to improve the productivity of other resources. By which, a firm could depend on it to effectively manage its all critical resources to achieve PF.

BDPA capability is one of the key organizational capabilities identified as the big competitiveness in the big data era with the help of CERP capability. The theoretical background identified three dimensions that support and reflect BDPA that features a higher-order construct, that is, data (D), MS and TS based on DCV. On top of that, CERP capability was frequently identified as a multi-dimensional construct with OF, PF and TF throughout our review and theoretical exploration. Further, we directly linked the relationship between BDPA and FP, CERP and FP (including MP and OP), respectively. Furthermore, we developed our hypotheses of these relationships (Figure 1).

3.1 Positive effect of cloud-based ERP on BDPA
Cloud-based ERP (CERP) links the internal departments of enterprise and the external enterprises across the supply chain, which shares a holistic view of all the data and
information that impacts enterprises’ performance. Data in CERP are becoming increasingly voluminous in size, form and velocity from business and transactional activities. CERP systems are increasingly exposed to big data wherein the data analytics take place in a short moment of time with huge amount of data in the various form. Big data enables CERP to have a contextual view regarding all the processes. Big data in CERP system originates from organizational process details, purchase histories, business interactions, web behavior and even social media (Sastry and Babu, 2013). According to Babu and Sastry (2014) and McAfee et al. (2012), big data analytics require common and great use of predictive analytics to unfold potential rules or the relationships to explore current data and historical facts and even visualize the hidden pattern for the decision management in high-volume and front-line operational decisions and future probabilities and trends in market (Matthias et al., 2017; Beatty and Williams, 2006). In other words, predictive analytics help firms to generate their weekly or even daily forecasts for material requirements planning for the sales goals in CERP system with the real-time analysis (John Lu, 2010; Babu and Sastry, 2014).

BDPA in CERP system provide forward-looking decisions to yield better operational effectiveness and more-informed market, improving the competitiveness of an enterprise among competitors and determining the likelihood of a future opportunity. It makes CERP system reflects not only what has happened, but also what is happening and will happen in the soon future for decision-making (Sistla and Babu, 2013; Salleh et al., 2012; Duan et al., 2013). From CERP system data, predictive analytics is the key branch of data mining and exploitation concerned with the optimization of the operational process and the prediction of future probabilities in the changing market (Snijders et al., 2012). Following what the literature indicated, we hypothesize the first on as:

**H1.** CERP service is positively related to the BDPA.

### 3.2 Positive effect of BDPA capability on FP

BDPA capability is widely acknowledged to play a vital role in improving FP in the changing markets (Akter et al., 2016; Wamba et al., 2017). The literature provides plenty of evidences of a relationship between BDPA and FP, including productivity optimization, profit, ROI and sales revenue maximization on OP (Schroeck et al., 2012; Gupta and George, 2016; Wamba et al., 2017); and market share and new markets, be more quicker in responding advantages of new products or services into market than other competitors on MP (Ramaswamy, 2013; Gupta and George, 2016; Wamba et al., 2017). More specifically, Srinivasan and Arunasalam (2013) concluded BDPA could enhance firms with a great benefit in healthcare of a firm by reducing the cost such as the fewer wastes and fraud and improving the quality of care such as operation and transaction safety and the efficacy of re-engineering and treatment. Thus, effectively exploiting these resources and capabilities into BDPA capability is phenomenal to maximize FP by facilitating the pervasive use of insights and support firms to achieve and sustain competitive advantages among competitors.

Drawing on DCV logic, we argued that superior FP (MP and OP) emerges from the effective exploitation of data, human such as MS and physical resources such as TS that are defined as valuable and so hard to replicate treasures of a firm. Validated by the BDPA literature, we argued that effective BDPA capability differentiates FP, creates firm precious competencies and advantages in the dynamically changing big data environment. Hence, we propose our second and third research hypotheses as:

**H2.** BDPA (BDPA) capability have a positive impact on the MP.

**H3.** BDPA (BDPA) capability have a positive impact on the OP.
3.3 Positive effect of cloud-based ERP on FP

CERP, as an integrated computer-based application, has been increasingly concentrated toward by most firms and hosted in cloud platform over the internet over the last few years due to its capability to handle high-volume data sets in the centralized database, flexibility, scalability, friendly-use, low cost of setup, improved accessibility, lower upfront and operating costs, rapid implementation, cost transparency, higher security standards and free trials (Beheshti, 2006; Duan et al., 2013; Salleh et al., 2012). By the record, CERP features managing information for decision-makers at right time, faster response, smoothing and routinizing flow of data and information across departments and enterprises, supporting third-party software, developing wide database regarding supplier, manufacturer and customer behaviors for future market competences (Johansson et al., 2015; Okezie et al., 2012; Appandairajan et al., 2012). Jain and Sharma (2016) found that adoption of CERP service had a critical impact on the day-to-day operations of a firm and simultaneously enhanced the efficiency of the operational processes. Yu et al. (2018) argued CERP provides enterprise effective interaction of cross-functional operation with the real-time integration of business operations.

In general, the authors found that CERP service helped most of the enterprises to lead more efficient business activities and new opportunities in the market armed by the seamless flow of integrated business operation and information (Dwayne Whitten et al., 2012; Battleson et al., 2016; Bruque Camara et al., 2015; Gupta and George, 2016; Gupta et al., 2018b). Hence, the literature leads to our fourth and fifth hypotheses as:

**H4.** CERP services have a positive impact on the MP of an organization.

**H5.** CERP services have a positive impact on the OP of an organization.

3.4 Moderating effects of organizational culture

Organizational culture correlates positively with the business performance that involves market share, product or service quality and responsivity, the source of competitive advantages in profit and sales (Jeble et al., 2018; Deshpandé et al., 1993). From the business perspective, organizational culture plays a vital role to support organizations in seeking competitive advantages for sustainable growth and thrive. As identified by many scholars (Hogan and Coote, 2014; Dubey, Gunasekaran, Childe, Papadopoulos, Hazen, Giannakis and Roubaud, 2017; Liu et al., 2010; Dubey, Gunasekaran, Childe, Papadopoulos, Hazen, Giannakis and Roubaud, 2017), organizational culture classified by control orientation (rational hand hierarchical culture) and flexible orientation (group and development culture) play their roles on different effects on organizational performance, respectively. Control-oriented culture commonly designs disciplines and regulations, focusing innovation initiatives (i.e. long-term strategies and goals), applying its core competencies and meeting business budgets. On the other hand, flexible-oriented culture enables firm’s creativity, empowerment and change vital for the exploration. Given flexibility-control would operate and originate from multiple levels. Control-oriented culture tends to descend from the top as the top management would come out with the strategy and goal, guidelines, disciplines and even constraints. By contrast, flexibility emerges in the model of bottom-up for most creative ideas or thoughts are popped up by the front-line subordinates. However, Khazanchi et al. (2007) argued that flexibility and control are not simple opposites and rather they contribute a more intricate paradox. Control is not the absence of flexibility and vice versa. In a nutshell, they would complement each other as the cross-fertilizations between the two orientations.

More specifically speaking, control-oriented culture emphasizes the predictability and efficiency, stability and productivity, cooperation, obeying disciplines and regulations and
focusing on innovation initiatives (Dubey, Gunasekaran, Childe, Papadopoulos, Hazen, Giannakis and Roubaud, 2017; Khazanchi et al., 2007). That conforms to CERP systems with BDPA capability that enables organization to ensure the business activities intelligible, interoperable, normative and legitimacy for normal and high-efficiency process operation, predictable and high-productive for more competences regarding market competitions (Deshpandé et al., 1993; Dubey, Gunasekaran, Childe, Papadopoulos, Luo, Wamba and Roubaud, 2017; Liu et al., 2010; Khazanchi et al., 2007). Firms with a high control orientation value efficiency, which is indeed a well-touted advantage for operation processing in CERP systems. In addition, focusing predictability features from control-oriented culture would be a high-level support for BDPA working, particularly in the market decision-making. Hence, the firm with a control orientation would be more likely to embrace CERP systems and BDPA capability. We argue that an enterprise with higher control orientation would be the one which shall probably focus on great operational and market benefit that is endowed with by such a seamless and timely BDPA capability on CERP platform and collaborate with internal and external partners, comparing with the lower control orientation competitors. Therefore, we suggest our hypotheses as:

\[ H_6. \] An organizational control orientation positively moderates the relationship between BDPA and MP.

\[ H_7. \] An organizational control orientation positively moderates the relationship between BDPA and OP.

\[ H_8. \] An organizational control orientation positively moderates the relationship between CERP and MP.

\[ H_9. \] An organizational control orientation positively moderates the relationship between CERP and OP.

By comparison, flexible orientation tends to refer to the group and development culture such as the creativity, innovation paradoxes, spontaneity, embracing changes and risk-taking (Liu et al., 2010; Dubey, Gunasekaran, Childe, Papadopoulos, Hazen, Giannakis and Roubaud, 2017). Consequently, a flexible orientation may support aligning organizational strategies and goals in the direction of the new market (new product or service) and unique operating systems. On top of that, firms with the flexibility-oriented culture tend to leverage various resources in CERP systems to develop distinct and competitive capabilities to differentiate from their competitors gain more advantages from heterogeneity across the market (Dubey, Gunasekaran, Childe, Papadopoulos, Luo, Wamba and Roubaud, 2017). In doing so, BDPA passively related to CERP would be enhanced with a more direct effect on the performance of a firm as flexible orientation enables firms more creative, open for embracing new opportunities or changes in the dynamically changing environment. Thus, we may argue that flexible orientation would positively moderate the relationships between CERP/BDPA and OP and MP respectively. Hence, we hypothesize:

\[ H_{10}. \] An organization's flexible orientation positively moderates the relationship between cloud ERP and OP.

\[ H_{11}. \] An organization's flexible orientation positively moderates the relationship between cloud ERP and MP.

\[ H_{12}. \] An organization's flexible orientation positively moderates the relationship between BDPA and OP.

\[ H_{13}. \] An organization's flexible orientation positively moderates the relationship between BDPA and MP.
4. Research methodology
The research methodology used in our study includes: developing a theoretical framework based on DCV theory to understand the role of cloud-based ERP services and BDPA on the performance of a firm and empirically validating this theoretical framework by employing SEM. In the third section, we have presented our theoretical framework as above. We would discuss the empirical validity next.

4.1 Construct operationalization
Appropriate scales from our literature review were employed to develop the survey instrument. The operationalization of each construct, measurement and derivation is shown in Table AI.

4.2 Data collection
Our study employed a survey-based approach. An online survey was carried out to gather data to test our hypotheses and then qualify the role of CERP and BDPA on the performance of a firm. We designed a questionnaire that was administered to after-sales support executive, AVP/VP/EVP, consultant, corporate finance executive/analyst, director/CEO/founder, engineer, manager/senior manager, and the sales/marketing executive (Table I) throughout four types of cloud services and the cloud solutions (Table II) with total 231 respondents. These 231 fully filled questionnaires were applied for our final data collection and analysis, after continually modifying and improving its clarity and appropriateness base on respondents’ inputs and feedbacks. Our target sample was those firms operating CERP with BDPA capability.

The key respondents we approached were not only involved directly in CERP but also interested in BDPA for they would be likely to provide the better response to our questionnaire. You could see the domain of our respondents and their work experiences

<table>
<thead>
<tr>
<th>Role in company/institution</th>
<th>Less than 10</th>
<th>10–50</th>
<th>50–300</th>
<th>300–500</th>
<th>500–1,000</th>
<th>More than 1,000</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>After-sales support executive</td>
<td>–</td>
<td>1</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>AVP/VP/EVP</td>
<td>–</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>Consultant</td>
<td>7</td>
<td>6</td>
<td>10</td>
<td>3</td>
<td>4</td>
<td>14</td>
<td>44</td>
</tr>
<tr>
<td>Corporate finance executive/Analyst</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>Director/CEO/Founder</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Engineer</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>30</td>
<td>46</td>
</tr>
<tr>
<td>Manager/Sr manager</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>11</td>
<td>51</td>
<td>84</td>
</tr>
<tr>
<td>Sales/Marketing executive</td>
<td>2</td>
<td>–</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>23</td>
<td>26</td>
<td>21</td>
<td>27</td>
<td>118</td>
<td>231</td>
</tr>
</tbody>
</table>

Table I. Role of employee in the company/institution and the number of employees

<table>
<thead>
<tr>
<th>Type of cloud services</th>
<th>Cloud service user</th>
<th>Cloud service provider</th>
<th>Cloud consultant or researcher</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software as a service (SaaS)</td>
<td>148</td>
<td>12</td>
<td>7</td>
<td>167</td>
</tr>
<tr>
<td>Platform as a service (PaaS)</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>Infrastructure as a service (IaaS)</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Internal cloud</td>
<td>27</td>
<td>1</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td>Total</td>
<td>201</td>
<td>16</td>
<td>14</td>
<td>231</td>
</tr>
</tbody>
</table>

Table II. Type of cloud services and the cloud solutions
in Table III, which would be well validated that our respondents ranged over a variety of fields. Table IV represents the age and qualification of the respondents. This helps to guarantee the quality of the better response from the respondents with different fields and stages for our subsequent data analysis as well. Hence, we could see data collection is an extremely important part for our later procedures regarding our study.

4.3 Data analysis and results

Our study puts efforts to assess the role of CERP and BDPA and even to explore the power of its combination on FP. Owing to the relationship between CERP and BPDA and its joint efforts for a firm were not examined in the previous research, there would be a few specific theoretical foundations for reference to anticipate their associations on FP. However, the extant studies indicate partial least squares (PLS) makes sense to estimate a general model for an exploratory research (Gupta and George, 2016; Henseler et al., 2014; Moshtari, 2016). And we supposed our study is a more exploratory research in nature. By similar argument, both almost chime in easily. Hence, we employed PLS–SEM performed by Warp PLS version 6.0 to test our model and research hypotheses reaping the benefits of effectiveness to explain and predict the target constructs from PLS–SEM (Hair et al., 2016; Akter et al., 2017).

Examining measurement model would be required before analyzing the PLS–SEM model. Variance inflation factors (VIF) is a way of identifying the risk of multicollinearity (Peng and Lai, 2012). VIF values of 5 or above would be considered to have a certain risk of multicollinearity, where we would improve our model. Our result Average block VIF showed a value of 4.966 (see Table V) and indicated that multicollinearity may not be a risk for our study. Although average block VIF values of 3 or below would be ideal fit value for PLS analysis, average path coefficient (APC) and average $R^2$ (ARS) would be mentioned regarding model fit indices for PLS as well. In Table V, APC (0.231, $p < 0.001$) and ARS (0.567, $p < 0.001$) show no evidence of fit problems. Hence, the results would be considered as a good fit to our model.
Apart from APC, ARS and AVIF, Sympon’s paradox ratio (SPR), $R^2$ contribution ratio (RSCR), statistical suppression ratio (SSR) and Nonlinear bivariate causality direction ratio (NLBCDR) are commonly used model fit indices for PLS. Concretely, these three indices would be regarded as causality assessment indices which aim to test our hypothesis are correct or not. SPR (0.715, the acceptable score), RSCR (0.913, the acceptable score and close to ideal score), SSR (1, the ideal score) and NLBCDR (0.785, the acceptable score) indicate that our theoretical model is appropriate for our study (see Table VI).

The result’s combined loadings and cross-loadings is also one part of the measurement model in PLS–SEM analysis and used with statistical tools Warp PLS version 6 to see our model is accepted or rejected (model fitness test). From the 231 fully filled questionnaire response, the combined loadings and cross-loading are developed to show the constructs of column and the indicators of row, as the indicators of reliability and validity of our theoretical model. Based on the results in the Appendix, it could group our model to qualify for the convergent validity into reflective indicators and significant factors. Loadings value of > 0.50 would be grouped into significant factor. $p$-value of < 0.05 are satisfied for reflective indicators. Therefore, our results (see the Appendix) indicate that most of the constructs are within or close to acceptable range in terms of reliability.

Cronbach’s $\alpha$ and composite reliability would be commonly used to test constructs reliability. In Table VII, the Cronbach’s $\alpha$ values are 0.966 for BDPA, 0.961 for CERP, 0.93 for MP, 0.943 for OP, 0.942 for CO and 0.914 for FO. In general, Cronbach’s $\alpha$ value of 0.7 and above is considered within the acceptable range (Gliner et al., 2001). We note that all the composite reliability is shown to arrive at the threshold value of 0.7 as well (Tellis et al., 2009). The average variances extracted (AVE) values satisfy the required value of > 0.5 (Hair et al., 2006) for latent variable. The greater value of $R^2$ coefficient means the better fitness to data. BDPA, MP and OP could be well explained by the constructs. All these results in Table VII could indicate the reliability of our constructs is acceptable.

Discriminant validity (irrelevance) tests whether an indicator that are not supposed to be related to a construct are actually unrelated. The acceptable discriminant validity is the case

<table>
<thead>
<tr>
<th>Average path coefficient (APC)</th>
<th>0.231, $p &lt; 0.001$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average $R^2$ (ARS)</td>
<td>0.567, $p &lt; 0.001$</td>
</tr>
<tr>
<td>Average block VIF (AVIF)</td>
<td>4.966, acceptable if $\leq 5$</td>
</tr>
</tbody>
</table>

**Table V.** Model fit and quality indices

<table>
<thead>
<tr>
<th>Sympon’s paradox ratio (SPR)</th>
<th>0.715, acceptable if $\geq 0.7$, ideally = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$ contribution ratio (RSCR)</td>
<td>0.913, acceptable if $\geq 0.9$, ideally = 1</td>
</tr>
<tr>
<td>Statistical suppression ratio (SSR)</td>
<td>1.000, acceptable if $\geq 0.7$</td>
</tr>
<tr>
<td>Nonlinear bivariate causality direction ratio (NLBCDR)</td>
<td>0.785, acceptable if $\geq 0.7$</td>
</tr>
</tbody>
</table>

**Table VI.** Causality assessment indices

<table>
<thead>
<tr>
<th>$R^2$ coefficients</th>
<th>BDPA</th>
<th>CERP</th>
<th>MP</th>
<th>OP</th>
<th>CO</th>
<th>FO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.631</td>
<td>–</td>
<td>0.429</td>
<td>0.64</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Adjusted $R^2$ coefficients</td>
<td>0.629</td>
<td>–</td>
<td>0.414</td>
<td>0.631</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Composite reliability coefficients</td>
<td>0.968</td>
<td>0.963</td>
<td>0.938</td>
<td>0.95</td>
<td>0.947</td>
<td>0.925</td>
</tr>
<tr>
<td>Cronbach’s $\alpha$ coefficients</td>
<td>0.966</td>
<td>0.961</td>
<td>0.93</td>
<td>0.943</td>
<td>0.942</td>
<td>0.914</td>
</tr>
<tr>
<td>Average variances extracted (AVE)</td>
<td>0.7</td>
<td>0.606</td>
<td>0.792</td>
<td>0.827</td>
<td>0.781</td>
<td>0.754</td>
</tr>
<tr>
<td>Variance inflation factors (VIF)</td>
<td>3.849</td>
<td>3.413</td>
<td>3.707</td>
<td>4.171</td>
<td>4.441</td>
<td>3.933</td>
</tr>
</tbody>
</table>

**Table VII.** Latent variable coefficients
where square roots of AVE values are greater than the construct correlations (Hair et al., 2006). Table VIII reveals the relationship among the constructs and the square root of AVEs. Results indicate that the constructs are positively correlated to each other, and square roots of AVEs are greater than the correlations values on the same column and row for that construct. Table VIII indicates that discriminant validity is acceptable for our paper.

Table IX indicates our hypotheses developed in Section 4 would be considered to be appropriate for our paper.

5. Discussion
5.1 Theoretical contribution

Our study is an early attempt to develop and explain the joint role of CERP and BDPA on the performance of a firm using the theoretical lens of DCV. From DCV’s perspective, our study extends the current understanding of concepts and the influence of BDPA and CERP

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$\beta$ and $p$-value</th>
<th>Supported or not supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1. Cloud-based ERP services is positively related to the big data predictive analytics</td>
<td>$\beta = 0.79$, $p &lt; 0.01$</td>
<td>Supported</td>
</tr>
<tr>
<td>H2. Big data predictive analytics has a positive impact on the market performance of an organization</td>
<td>$\beta = 0.51$, $p &lt; 0.01$</td>
<td>Supported</td>
</tr>
<tr>
<td>H3. Big data predictive analytics has a positive impact on the operational performance of an organization</td>
<td>$\beta = 0.48$, $p &lt; 0.01$</td>
<td>Supported</td>
</tr>
<tr>
<td>H4. Cloud-based ERP services has a positive impact on the market performance of an organization</td>
<td>$\beta = 0.28$, $p &lt; 0.01$</td>
<td>Supported</td>
</tr>
<tr>
<td>H5. Cloud-based ERP services has a positive impact on the operational performance of an organization</td>
<td>$\beta = 0.32$, $p &lt; 0.01$</td>
<td>Supported</td>
</tr>
<tr>
<td>H6. An organization’s control orientation positively moderates the relationship between BDPA and market performance</td>
<td>$\beta = 0.18$, $p &lt; 0.01$</td>
<td>Supported</td>
</tr>
<tr>
<td>H7. An organization’s control orientation positively moderates the relationship between BDPA and operational performance</td>
<td>$\beta = 0.01$, $p = 0.42$</td>
<td>Nota supported</td>
</tr>
<tr>
<td>H8. An organization’s control orientation positively moderates the relationship between cloud ERP and market performance</td>
<td>$\beta = 0.02$, $p = 0.36$</td>
<td>Nota supported</td>
</tr>
<tr>
<td>H9. An organization’s control orientation positively moderates the relationship between cloud ERP and operational performance</td>
<td>$\beta = 0.10$, $p = 0.06$</td>
<td>Not supported</td>
</tr>
<tr>
<td>H10. An organization’s flexible orientation positively moderates the relationship between cloud ERP and operational performance</td>
<td>$\beta = 0.02$, $p = 0.36$</td>
<td>Not supported</td>
</tr>
<tr>
<td>H11. An organization’s flexible orientation positively moderates the relationship between cloud ERP and market performance</td>
<td>$\beta = 0.02$, $p = 0.39$</td>
<td>Not supported</td>
</tr>
<tr>
<td>H12. An organization’s flexible orientation positively moderates the relationship between BDPA and operational performance</td>
<td>$\beta = 0.02$, $p = 0.40$</td>
<td>Not supported</td>
</tr>
<tr>
<td>H13. An organization’s flexible orientation positively moderates the relationship between BDPA and market performance</td>
<td>$\beta = 0.23$, $p &lt; 0.01$</td>
<td>Supported</td>
</tr>
</tbody>
</table>

Table VIII.
Correlations among latent variables with square root of AVEs

<table>
<thead>
<tr>
<th></th>
<th>BDPA</th>
<th>CERP</th>
<th>MP</th>
<th>OP</th>
<th>CO</th>
<th>FO</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDPA</td>
<td>0.837</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERP</td>
<td>0.794</td>
<td>0.778</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MP</td>
<td>0.747</td>
<td>0.7</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP</td>
<td>0.765</td>
<td>0.732</td>
<td>0.821</td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>0.699</td>
<td>0.7</td>
<td>0.706</td>
<td>0.72</td>
<td>0.884</td>
<td></td>
</tr>
<tr>
<td>FO</td>
<td>0.723</td>
<td>0.693</td>
<td>0.695</td>
<td>0.707</td>
<td>0.835</td>
<td>0.869</td>
</tr>
</tbody>
</table>

Table IX.
Results of hypotheses testing
as organizational capabilities on FP (Teece et al., 1997; Smith et al., 2014; Lawson and Samson, 2001; Kraaijenbrink et al., 2010). There exists remarkable contributions made by practitioners of CERP or BDPA to understand the effects of CERP or BDPA solely, but not from the prospective of integrated association of the two (Battleson et al., 2016; Chen et al., 2015; Akter et al., 2016; Wamba et al., 2015). We addressed the shortcoming by highlighting the joint effects of CERP and BDPA to improve the PF. Further, our empirical results using survey data from 231 executive-level technology respondents suggest that: for one, CERP is positively related to BDPA (H1), which supports our study on the joint effects of both; for another, BDPA directly affect MP and OP (H2 and H3) in dynamically changing environment, the same in CERP aspects (H4 and H5). Therefore, we could say our study further extend the extant CERP and BDPA literature regarding bringing organizations potential and sustainable development.

Our study also asserted the organizational cultures (i.e. control-oriented and flexibility-oriented) were not positively moderates the relationship between BDPA and operation performance (H7 and H12), between CERP and MP/OP (i.e. H8/H9 and H10/H11). The results are contrary to our expectations, but we supposed the relationship between organization culture and BDPA and CERP, respectively would happen to be not consistent to the previous literature (Schroeck et al., 2012; Fosso Wamba et al., 2017; Liu et al., 2010; Dubey, Gunasekaran, Childe, Papadopoulos, Luo, Wamba and Roubaud, 2017). It would also be an interesting question that is worth pursing further for future research scholars.

Finally, another important contribution of our study is to develop a theoretically grounded construct of organization survey instrument to measure the organization's ability with CERP services and BDPA capability that is different from digital capabilities (i.e. IT capability). The constructs and its measures would give references to further researchers devoted to studying CERP and BDPA on the performance of an organization in terms of further the emerging research stream in that regard.

5.2 Managerial implications
Our study yields interesting insights for practice in business organizations for improving firm's competencies and competitiveness in a dynamically changing environment. Based on our current study highlighting the positive influence of CERP and BDPA on the performance of a firm, it attempts to enlighten business organizations that catching competitiveness in a changing market is not only about making hordes of investment, employing more talents with sophisticated skills and techniques, but also having access to CERP services as a well-established and cost-effective platform for data and information generation, sharing, processing and subsequent predictive analysis, BDPA capability where potential insights extracted from data and analysis acted upon and data-driven organizational culture (i.e. control-oriented and flexibility-oriented) where employees from top to ground have awareness to cultivate and process data-specific technologies and skills.

In addition to this, our study contained empirical validation of the role of CERP and BDPA on FP. It has advanced the understanding for top management and executives regarding the positive impacts in MP and OP from CERP and BDPA to enhance a firm's competitiveness and sustainability in the dynamically changing environment. Our study suggested that business firms implement CERP service and invest in very talent techniques and knowledge learning and sharing culture to build BDPA capability, which offers huge benefits to gain competitive advantages among competitors.

5.3 Limitations and directions for further research
Notwithstanding the insights on the positive impacts of CERP and BDPA for organizational performance for researchers and industry practitioners, some limitations and future research directions would be outlined. First, our study is the investigation of the role of

Role of CERP and big data on firm performance

5.2 Managerial implications

5.3 Limitations and directions for further research

1873
CERP and BDPA as organizational capabilities that impact on FP and bring competitive advantages for a firm. The final destination, we hope interested business organizations could understand and build CERP and BDPA capabilities for their firms. In such an attempt, further research may need to be explored through investigating the role of institutional pressures on implementing CERP systems and their commitment toward developing BDPA capability, which could help to make it possible to employ CERP systems with BDPA capability in firms. In this vein, our research concept is not only one scientific idea, but one kind of advanced firm operation type.

Second, both sets of data needed for the empirical validation in our study focused on the organizations from India. Furthermore, the study conducted the survey-based empirical validation at one point in time from 231 respondents. Hence, our study would be expanded by involving a broader sample outside of India, avoiding the limitation of our understanding of the role of CERP and BDPA on FP caused by gathering data from single source at a single point in time. It would be interesting for our study to see if the country-level differences would influence these relationships between CERP and FP, BDPA and FP, and CERP and BDPA.

Finally, our study relies on the survey-based approach for empirical validation. To advance better insights into CERP and BDPA capabilities, maybe the way of a mixed form approach could do better, such as telephone interview, semi-structured interviews with respondents. By which, the relationships between these constructs in our model could be further understood and lead to better empirical investigation.

6. Conclusion
This work developed a theoretical framework based on DCV theory to understand the role of CERP and BDPA on the performance of a firm. Drawing broadly on the DCV theory, we have conceptualized CERP and BDPA as organizational capabilities and tested the relationship between CERP and BDPA and their respective impacts on FP under the moderating effects of organizational culture. Through our empirical validation by employing SEM, it turned out that CERP which is positively related to BDPA has positive impacts on FP, and BDPA has the same positive impacts on FP, although the high or low control and flexible orientation do not positively moderate the impacts of BDPA on OP, CERP on MP and OP. By which, it supports to a certain degree to advance the understanding for business organizations to consider and build CERP and BDPA capabilities. Our study would be first theoretical framework to investigate and explain the joint impacts of CERP and BDPA on FP. And finally, we presented the empirical evidence that CERP and BDPA would be predictors of the organizational performance by developing a survey-based instrument. However, further research based on our limitations part would be explored to arrive to our final goal regarding convincing organizations to build CERP and BDAP capabilities.

References


### Appendix

<table>
<thead>
<tr>
<th>Latent variable</th>
<th>Indicator</th>
<th>Measurement constructs</th>
<th>Journal paper considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organizational factors (OF)</td>
<td>OF1</td>
<td>Strategic goals and objectives</td>
<td>Gupta and Misra (2016)</td>
</tr>
<tr>
<td></td>
<td>OF2</td>
<td>Communication</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OF3</td>
<td>Implementation strategy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OF4</td>
<td>Business process re-engineering</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OF5</td>
<td>Project management</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OF6</td>
<td>Project budget</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OF7</td>
<td>Organization resistance</td>
<td></td>
</tr>
<tr>
<td>People factors (PF)</td>
<td>PF1</td>
<td>User involvement</td>
<td>Gupta and Misra (2016)</td>
</tr>
<tr>
<td></td>
<td>PF2</td>
<td>Selection of vendor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PF3</td>
<td>Project team</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PF4</td>
<td>Top management support</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PF5</td>
<td>Training of user</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PF6</td>
<td>Trust on vendor</td>
<td></td>
</tr>
<tr>
<td>Technological factors (TF)</td>
<td>TF1</td>
<td>Selection of ERP package</td>
<td>Gupta and Misra (2016)</td>
</tr>
<tr>
<td></td>
<td>TF2</td>
<td>IT infrastructure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TF3</td>
<td>Data integrity and system testing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TF4</td>
<td>Functionality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D1</td>
<td>Integrate data from multiple internal sources into a data warehouse for easy access</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>Integrate external data with internal to facilitate high-value analysis of business environment</td>
<td></td>
</tr>
<tr>
<td>Managerial skills (MS)</td>
<td>MS1</td>
<td>Big data analytics managers are able to work with functional managers, suppliers and customers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS2</td>
<td>Big data analytics managers are able to coordinate big data-related activities</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS3</td>
<td>Big data analytics managers are able to anticipate the future business needs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS4</td>
<td>Big data analytics managers have a good sense of where to apply big data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MS5</td>
<td>Big data analytics managers are able to understand and evaluate the output extracted from big data</td>
<td></td>
</tr>
<tr>
<td>Technical skills (TS)</td>
<td>TS1</td>
<td>Big data analytics training to employees</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TS2</td>
<td>Hire new employees that already have the big data analytics skills</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TS3</td>
<td>Big data analytics staff has the right skills to accomplish their jobs successfully</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TS4</td>
<td>Big data analytics staff has suitable education to fulfill their jobs</td>
<td></td>
</tr>
</tbody>
</table>

**Table AI.**
Operationalization of Constructs

(continued)
<table>
<thead>
<tr>
<th>Latent variable</th>
<th>Indicator</th>
<th>Measurement constructs</th>
<th>Journal paper considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS</td>
<td>TS5</td>
<td>Big data analytics staff holds suitable work experience to accomplish their jobs successfully</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CO2</td>
<td>Leveraging core competencies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CO3</td>
<td>Profile valuation</td>
<td>Dubey, Gunasekaran, Childe, Papadopoulos, Luo, Wamba and Roubaud (2017)</td>
</tr>
<tr>
<td></td>
<td>CO4</td>
<td>Value–practice interactions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CO5</td>
<td>Value congruence</td>
<td></td>
</tr>
<tr>
<td>Flexible orientation (FO)</td>
<td>FO1</td>
<td>Innovation paradoxes</td>
<td>Deshpande et al. (1993), Khazanchi et al. (2007), Liu et al. (2010); Dubey, Gunasekaran, Childe, Papadopoulos, Hazen, Giannakis and Roubaud (2017), Dubey, Gunasekaran, Childe, Papadopoulos, Luo, Wamba and Roubaud (2017)</td>
</tr>
<tr>
<td></td>
<td>FO2</td>
<td>Creativity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FO3</td>
<td>Spontaneity</td>
<td>Papadopoulos, Luo, Wamba and Roubaud (2017)</td>
</tr>
<tr>
<td></td>
<td>FO4</td>
<td>Risk-taking</td>
<td></td>
</tr>
<tr>
<td>Firm performance (FP)</td>
<td>MP1</td>
<td>Exploring new markets more quickly than competitors</td>
<td>Gupta and George (2016)</td>
</tr>
<tr>
<td></td>
<td>MP2</td>
<td>Introducing new products or services into the market faster than competitors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MP3</td>
<td>Success rate of new products or services has been higher than competitors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MP4</td>
<td>Market share has exceeded that of competitors</td>
<td></td>
</tr>
<tr>
<td>Operational performance (OP)</td>
<td>OP1</td>
<td>Productivity has exceeded compared to competitors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OP2</td>
<td>Profit rate has exceeded compared to competitors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OP3</td>
<td>Return on investment (ROI) has exceeded compared to competitors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OP4</td>
<td>Sales revenue has exceeded compared to competitors</td>
<td></td>
</tr>
</tbody>
</table>

Table AI.
MD
57,8

1882

Table AII.
Combined loadings
and cross-loadings

D1
D2
D3
MS1
MS2
MS3
MS4
MS5
TS1
TS2
TS3
TS4
TS5
OF1
OF2
0F3
OF4
OF5
OF6
OF7
PF1
PF2
PF3
PF4
PF5
PF6
TF1
TF2
TF3
TF4
MP1
MP2
MP3
MP4
OP1
OP2
OP3
OP4
CO1
CO2
CO3
CO4
CO5
FO1
FO2
FO3
FO4

BDPA

CERP

MP

OP

CO

FO

SE

p-value

0.664
0.709
0.739
0.878
0.902
0.907
0.904
0.913
0.816
0.814
0.87
0.858
0.855
−0.256
−0.143
0.195
−0.117
−0.261
−0.096
−0.11
−0.007
0.075
0.164
−0.152
0.064
0.103
0.037
0.034
0.281
0.151
0.024
0.058
−0.035
−0.147
0.046
−0.075
−0.007
−0.038
−0.13
−0.038
−0.054
0.111
0.043
0.014
−0.05
−0.049
−0.147

0.589
0.407
0.388
−0.004
0.061
0.071
0.025
0.019
−0.182
−0.31
−0.299
−0.135
−0.28
0.77
0.795
0.78
0.794
0.816
0.775
0.802
0.776
0.804
0.805
0.803
0.806
0.807
0.754
0.629
0.747
0.75
−0.072
−0.012
−0.071
−0.053
0.178
−0.028
0.003
−0.076
−0.047
−0.062
−0.075
−0.02
0.087
−0.079
−0.053
0.045
−0.056

0.094
−0.098
−0.049
−0.049
−0.2
−0.093
0.081
−0.008
−0.04
−0.032
−0.014
−0.144
−0.074
−0.215
0.002
0.217
0.068
0.047
0.177
0.166
0.091
−0.062
−0.202
−0.386
−0.199
−0.208
−0.184
0.209
0.048
−0.068
0.885
0.916
0.92
0.836
−0.296
−0.013
−0.073
0.008
0.047
0.08
−0.074
−0.115
−0.24
0.066
−0.188
−0.056
−0.012

−0.405
0.113
0.003
0.021
0.203
−0.01
−0.14
−0.031
0.015
0.084
0.049
−0.084
0.026
0.254
0.063
−0.28
0.051
0.04
−0.206
−0.197
−0.255
0
0.016
0.297
0.134
−0.078
−0.059
−0.234
−0.115
−0.133
−0.01
−0.015
−0.077
0.2
0.836
0.937
0.932
0.93
−0.028
−0.026
0.081
−0.13
0.061
−0.008
0.092
−0.078
0.059

−0.103
−0.54
−0.228
−0.183
−0.294
−0.259
−0.214
−0.135
0.236
0.239
0.407
0.282
0.477
−0.017
−0.163
−0.015
−0.001
−0.004
0.136
0.077
0.109
−0.071
0.037
0.001
−0.157
−0.279
−0.048
0.034
−0.24
0.072
−0.04
−0.246
0.081
0.232
−0.41
−0.07
0.01
0.036
0.879
0.889
0.895
0.889
0.867
−0.067
−0.172
0.026
−0.125

−0.077
0.108
−0.142
0.047
0.188
0.155
0.171
0.085
−0.155
−0.144
−0.325
−0.184
−0.292
−0.01
0.014
−0.124
−0.106
0.037
−0.162
−0.061
−0.163
0.017
−0.117
0.046
−0.013
0.169
0.168
−0.115
0.124
0.105
−0.05
0.055
−0.105
−0.179
0.258
0.016
−0.123
−0.105
0.133
−0.087
−0.163
−0.048
−0.159
0.817
0.922
0.891
0.841

0.058
0.058
0.058
0.056
0.056
0.056
0.056
0.056
0.057
0.057
0.056
0.056
0.056
0.057
0.057
0.057
0.057
0.057
0.057
0.057
0.057
0.057
0.057
0.057
0.057
0.057
0.057
0.059
0.058
0.058
0.056
0.056
0.056
0.057
0.057
0.056
0.056
0.056
0.056
0.056
0.056
0.056
0.056
0.057
0.056
0.056
0.057

o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001
o0.001

Note: Loadings are un-rotated and cross-loadings are oblique-rotated, both after separate Kaiser normalizations

Corresponding author
Zongwei Luo can be contacted at: luozw@sustc.edu.cn

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com


Interplay between information systems and environmental management in ISO 14001-certified companies
Implications for future research on big data

Paula De Camargo Fiorini
Department of Production Engineering,
Sao Paulo State University (UNESP), Bauru, Brazil
Charbel Jose Chiappetta Jabbour and
Ana Beatriz Lopes de Sousa Jabbour
Montpellier Business School, Montpellier Research in Management,
Montpellier, France
Nelson Oliveira Stefanelli
Business School, University of São Paulo, Ribeirão Preto, Brazil, and
Yudi Fernando
Faculty of Industrial Management,
Universiti Malaysia Pahang, Kuantan, Malaysia and
Management Department, Bina Nusantara University, Jakarta Barat, Indonesia

Abstract
Purpose – The purpose of this paper is to identify the contributions of information systems (IS) for the evolutionary process of corporate environmental management by highlighting implications for big data research.
Design/methodology/approach – The authors conducted two case studies with Brazilian enterprises certified by ISO 14001, by conducting interviews, document analysis and direct observation. Implications for a research agenda on big data are also presented.
Findings – As results, the authors present the identification of the main contributions of IS for the evolution of environmental management in the studied cases. The authors found that advanced stage regarding IS may be considered a factor that implies a more effective environmental management.
Originality/value – The main contribution of this research consists of the presentation of a framework that identifies the support of IS for corporate environmental practices. By confirming the relation between IS and maturity levels of environmental management, the authors highlight that application of big data has the potential of boosting the relation between IS and corporate environmental management.

Keywords Information systems, ISO 14001, Big data, Sustainable operations,
Corporate environmental management

Paper type Research paper

1. Introduction
Corporate sustainability management has been a subject of high relevance in public and private agendas in different countries, being recognized as an essential factor for the success of organizations (Dubey et al., 2017). It concerns the inclusion of environmental, social and economic considerations in business activities (Yang et al., 2018). To properly accomplish corporate sustainability management, especially the environmental dimension, companies

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.
need to handle large volumes of data (Seles et al., 2018). In this sense, decision-support tools are needed to provide more organized, understandable, and accurate information in order to predict and prevent unsustainable practices (O’Rourke, 2014).

In this context, information systems (IS) are vitally relevant for managing, filtering and systematizing useful information for the implementation and maintenance of environmental management systems (EMSS) (Fiorini and Jabbour, 2017). According to Laudon and Laudon (2017), IS have played a fundamental role in organizations, as they are among the most important tools to be used to achieve high levels of efficiency and productivity in operations. Through them, decision makers have easy access to the information of the entire organizational set, which significantly supports decision making.

Organizations are increasingly concerned with monitoring information, which is related to how to identify and manage social, ethical and environmental risks, and explain how these risks may be important in the short and long term (Walker et al., 2007). Based on this understanding, IS are fundamental tools for environmental management (Schimak, 2005). As the volume of environmental data increases, it is possible to point out that big data (Mishra et al., 2018; Fosso Wamba et al., 2018; Prescott, 2014) will profoundly change the IS provide support for corporate environmental management (Seles et al., 2018; Dubey et al., 2017).

However, despite all the relevance of environmental management and IS in the organizational context, the literature regarding their relationship is still considered scarce (Fiorini and Jabbour, 2017). Indeed, the relation between advanced information management and sustainable production has only recently been explored (Dubey et al., 2018). According to Ryoo and Koo (2013), IS offer an excellent opportunity to contribute to environmental sustainability in the organizational context.

The adoption of environmental actions requires new data regarding impacts and information on causes and effects, as well as sharing knowledge about the effectiveness of environmental practices. Given the situation, IS are an important but inadequately understood tool for organizations in search of environmental sustainability, capable of making new practices and processes available, supporting reactions and evaluating results (Melville, 2010).

Confirming the proposition above, Jenkin et al. (2011) have suggested that there are still many gaps representing the opportunities for research focused on the relationship between environmental sustainability and IS. Melville (2010) reinforces that few studies on environmental sustainability incorporate the IS perspective. Also, according to Sarkis et al. (2013), subjects such as the contribution of IS to internal and external environmental awareness to organizations, as well as the roles of IS in environmental-oriented business practices have a lack of studies.

Following the same line, existing research on information technology and IS in conjunction with sustainability have concentrated their efforts on the use of these to reduce the environmental impacts and energy consumption in corporate IT systems (see Cai et al., 2013; Watson et al., 2010). For Dao et al. (2011), research on the value of IT sustainability for environmental sustainability needs to go beyond operational initiatives to reduce energy consumption.

In this paper, we explore the potential of IS to support environmental practices (organizational, operational and communicational) as established by González-Benito and González-Benito (2006). We present case studies on the relation between IS and the evolution of corporate environmental management in ISO 14001-certified companies in an emerging economy: Brazil. From the research results, we provide unique insights on the future of big data in the relation IS-environmental management. Therefore, the specific objectives associated with the research are:

1. to understand the contributions of IS to environmental management practices in the cases studied;

2. to identify barriers and opportunities to the support of IS to environmental management; and
(3) to draw a research agenda on the potential role that big data will play in the relation IS-corporate environmental management.

Following, we present a conceptual synthesis on the topics needed to understand the study: Environmental management in organizations, IS and the emergence of big data, and considerations about the relationship between IS and environmental management. Then, we report the research methodology employed in the conduction of the study and the main results. Finally, the sections of discussions, conclusions and suggestions for future research are presented.

2. Conceptual background

2.1 Environmental management in organizations

In 1972, with the first United Nations Conference on the Human Environment – Stockholm Conference – the environment and especially the relationship between it and the organizations has become a relevant topic for public policy and for business strategy. As a direct result of this Conference, the United Nations Environment Programme was created. The World Commission on Environment and Development and a large part of industrialized countries designed Environmental ministries, secretariats and agencies. In 1987, the World Commission on Environment published a report titled Our Common Future, also known as the Brundtland report, a milestone in the history of environmental management, dedicating the concept of sustainable development and clearly establishing the important role that companies should have in environmental management. As a result, the United Nations Conference on Environment and Development was scheduled for 1992, based in Rio de Janeiro, highlighting the importance of environmental management at intergovernmental level.

In this context, it is clear that the environment has become a vital element in establishing the new paradigms of industrial competition and has therefore emerged as an important field for research and business practice in the last decades (Dao et al., 2011; Kolk and Mauser, 2002).

The globalization of environmental problems is an undeniable fact and, consequently, environmental management practices change the image of organizations, becoming priorities for the future stages of business management and financial investments of the organizations. As part of an environmentally changing society, the companies have a large share of responsibility for achieving sustainable development (Jabbour and Santos, 2006).

In this way, organizations are increasingly interested in the incorporation of the environmental variable, because of the increasing awareness in political and social terms, about the need for environmental preservation. Customers and consumers have become demanding about the environmental aspect of the products.

According to Christie et al. (1995), environmental management is a set of techniques and disciplines that drive companies in adopting a cleaner and action-focused production for the prevention of losses and pollution. The literature on this topic has progressed in a relevant way worldwide. Authors from all over the world discuss this theme, expressing the growing consideration that the environment is no longer an obstacle or a restriction, but a new business opportunity. Evans et al. (2017) affirm that environmental management has become an important tool for organizational innovation and competitiveness.

González-Benito and González-Benito (2006) distinguish three categories of environmental practices, exercised by environmentally proactive organizations – understood as the voluntary implementation of practices, and initiatives aimed to improve the environmental performance. These categories are:

1. Planning and organizational practices: they refer to the implementation of an EMS, i.e. the definition of an environmental policy, the development of procedures for establishing environmental objectives, the selection and implementation of environmental practices. These practices do not reduce environmental damage
but establish mechanisms that lead to a company’s advancement regarding this aspect.

(2) Operational practices: they imply changes in the production system and operations, and can be classified into two groups: related to the product and related to the operational processes. The first includes practices focused on design and development of environmentally-conscious products, such as design for the elimination of hazardous and polluting materials in products. The second group focuses on the development and implementation of environmentally-conscious production methods and processes.

(3) Communicational practices: they aim to communicate and disseminate environmental actions to stakeholders in the company. These practices seek business objectives and the establishment of relations with stakeholders. They are not focused on improving environmental performance, but disseminating it.

As presented above, one of the environmental practices in the organizational context is the development of an EMS. Thus, among the multiple instruments that aim to boost the business environmental management process, the EMSs are highlighted.

2.2 EMS based on ISO 14001

Given the increasing pressure suffered by organizations for better administration of the environmental issue, an increase in the adoption of EMS is observed (Mazzi et al., 2016; Fryxell and Szeto, 2002). The term EMS concerns all organizational actions carried out in a systematic way to monitor the environmental impacts of their activities and to manage issues relevant to the green dimension (Elefsiniotis and Wareham, 2005).

In addition to the contribution to social responsibility and the creation of the circumstances for compliance with the legislation, these systems make it possible to identify opportunities to reduce the use of materials and energy, as well as to improve the efficiency of the processes (Chan and Wong, 2006).

EMS legitimized with ISO 14001 are highly demanded, as this certification provides the most recognized EMS structure in the world (Mazzi et al., 2016). It provides tools for an organization to better manage the environmental impacts of its activities, improving their environmental performance and bringing a number of operational, financial and corporate benefits (Prajogo et al., 2014; Gavrornski et al., 2008).

In Brazil, the number of companies that develop environmental management based on standard ISO 14001 has been increasing every year. Ecological awareness is paving the way for the development of new business opportunities and thus facilitating the inclusion of Brazilian companies in the international market (Silva and Medeiros, 2004).

The standard ISO 14001 lays down requirements for management of EMSs without defining the form and degree that they should have or achieve, thus allowing companies to develop their own solutions to meet the requirements of the norm (Mazzi et al., 2016). This gives it a universal character because they can be adapted by companies from any region and all sizes (ISO, 2015).

As the ISO 14001 itself highlights, it must be appropriate to the company’s activities, products and processes. It must include the commitment of all with continuous improvement and pollution prevention, reflect the values and principles of the organization, comply with existing legislation, to ensure the provision of the necessary infrastructure, to be documented and disseminated to the entire organization and should be available to the external public (ISO, 2015).

The organizational environmental management, then, consists of the implementation of programs focused on the development of environmentally conscious technologies
and products, which seek to meet legal issues, but also to take advantage of business opportunities, at the same time it improves the institutional image. The ISO 14001 standard, then embedded in the presented panorama, is worth as a legitimate model for organizations to formalize environmental management in their operations.

2.3 IS in organizations: the emergence of big data

There is a consensus that in the industrial society, whose economy assumes global trends, the information has become a precious capital, equating to the production, material and financial resources. To be successful, IS management should rely on understanding organizations’ context and macro-environment. The relevance of truly aligning organizations’ objectives and the macro-environment has been recognized as a key component of the contemporary management (Gaur et al., 2011; Mukherjee et al., 2013).

According to Peter Drucker (1993), the activities that occupy the central place of the organizations are no longer those that aim to produce or distribute objects, but those that produce and distribute information and knowledge. Thus, in an information resource-based economy, McGee and Prusak (1994) claim that competition among organizations is based on their ability to acquire, treat, interpret and use information effectively.

In view of this scenario, it is remarkable how important it is to develop information management effectively and to highlight the fundamental role that information technology exerts in this process. Thus, for Rezende (2002), it will be the skill with which the organization collects, organizes, analyzes and implements changes from information, integrating them into the process of continuous improvement of its activities, which will determine its excellence.

In this context of relevant emphasis on the approach of information, IS are presented as a tool to support organizations, mainly to their managers, who pass on information from their organizations as a whole, broadening their perceptions about them.

For O’Brien and Marakas (2010), IS can be understood as an organized set of people, hardware, software, communications networks and data resources that collects, transforms and disseminates information into an organization. Laudon and Laudon (2017) define IS, technically, as a set of interrelated components that collect, process, store and distribute information to support decision making, the coordination and control of an organization. Thus, the characteristics of the IS demonstrate their main ability, to provide information for control and agility in decision making.

There are several types of IS that manipulate a variety of information at different organizational levels. However, in view of the large variations that the IS have in the present day, it can be said in general that it does not matter the nomenclature attributed, since its objectives always reside in generating information about the activities of organization, which assist managers to become informed and sensible in making decisions about their operational and strategic activities (Smith, 1999).

The growth of integrated IS provides the managers with extraordinarily detailed data about the activities of the organization, which broadens the speed and accuracy of the decisions. Organizations today use IS to achieve six main objectives:

1. operational excellence;
2. new products and business models;
3. close relationship with customers and suppliers;
4. better decision making;
5. competitive advantage; and
6. daily survival.
An information system should not be understood as equivalent only to IT, but as the set of three types of elements – technology, organization and people – which together form a technical partner system (Laudon and Laudon, 2017). This social technical characteristic of an IS implies in several challenges to its process of implantation. In this scenario, the technical, social and human elements interact within a context, from which unpredictable conditions arise resulting in changes in technology and people (Wyrwicka et al., 2018; Caldeira and Ward, 2003).

Still, the implementation of IS is an important process within the organization because if successful, potential benefits can be achieved such as increase in sales, profitability, productivity, improvement in the process of decision making and safer competitive positions on the market (Thong, 2001).

After all, it is impossible to talk about information management in the twenty-first century without mentioning big data. Big data have been considered a major revolution in the way companies collect, process and deal with large data (McAfee and Brynjolfsson, 2012), impacting decision-makers capabilities (Frisk and Bannister, 2017; Davenport, 2014). Akter et al. (2016) mention that the company’s technological systems and capabilities are key dimensions for better exploitation of big data in several management functions. The definition of big data implies an understanding on the “Vs” of big data, such as volume, variety, and velocity (Fosso Wamba et al., 2015). Big data have impacted a variety of business functions, including emergent topics in sustainability (El-Kassar and Singh, 2018; Jabbour et al., 2017).

In this research and based on latest findings (Song et al., 2017; Papadopoulos et al., 2017) this work considers that the relation between IS and corporate environmental management can have implications for the research agenda on big data. But before understanding the implications of IS-corporate environmental management, it is important to discuss the relationship between IS and green management further, as below.

2.4 Considerations on the relationship between IS and environmental management
The highlighted relevance of information in all organizational levels is indispensable to environmental management. Based on some principles defended by Gaur and Kumar (2018), it is possible to systematize the main works on the relation of IS and environmental management as shown in Table I.

Organizations have a constant need to make decisions related to environmental issues. These decisions, in turn, involve a large number of information and the participation of many people (Dionysio and Santos, 2008).

According to Boiral (2006), issues related to political, social and scientific or technical monitoring, which can affect the activities of an organization, are interdependent and they require a multidisciplinary approach, integrating a wide variety of information. The complexity of this type of information has led some organizations to cooperate in the exchange of environmental information. This group of initiatives we find for example EIATRACK, a portal funded by the industry’s leading electronic products organizations to ensure environmental information from the major countries of the world, which are crucial to adapting the projects of products and increase the scope of environmental regulations.

The IS strongly represent an instrument for environmental sustainability, but they are not properly clarified and exploited. They represent a valuable source of possible solutions to environmental issues. IS are significantly influential in creating innovative environmental strategies and play a key role in reducing environmental pollution, developing technology solutions and creating sustainable business opportunities (Lagumdzija et al., 2012).

Still under the prism of sustainability, IS can enable organizations to standardize, monitor, capture and use data and metadata that help evaluate the economic, environmental and social impacts of the organization’s activities (Melville, 2010). According to Dao et al. (2011),
close collaborations and exchange of information within and between organizations through IS are fundamental to the development of sustainable capacities.

In this scenario, an enterprise resource planning (ERP) system or a database that incorporates environmental aspects can, for example, help organizations collect and share data about their environmental performance (Chen et al., 2008). Web portals would make possible to provide transparent information on the social and environmental impacts of products or to enable collaboration with external stakeholders, such as potential partners and local communities (Dao et al., 2011).

In summary, substantially, to increase its environmental capacity, an organization should align environmental practices with IS (Dao et al., 2011). This alignment, for organizational structures, ensures the support of IS to reach their environmental objectives (Fiorini and Jabbour, 2017; Ryoo and Koo, 2013).

In the various phases of an EMS, implementation and operation, verification and analysis, different types of information are requested, on human, financial, document-related resources, results of periodic evaluations, audits, among others. For example, interorganizational information exchange can facilitate the efficient flows of knowledge and experience required by an EMS (Prajogo et al., 2014). IS can play a central role as tools for improving knowledge sharing and sustainability indicators (Bengtsson and Agerfalk, 2011).
Technologies, in particular information technologies, represent the source of potential solutions to environmental problems. Higher education institutions, IS associations and systems-oriented journals should use their significant position as an influence so that academic leaders and students face the challenge of investigating the area of information systems and sustainability (Lagumdzija et al., 2012).

There are still many open challenges to examine in this area that deserve the attention of the research community (Melville, 2010). Watson et al. (2010) establish general eco-goals, which drive the actions and thoughts in the area of sustainability and can receive aid from IS making the investigation of this theme strategically important.

In general, considering the importance of information as a structuring factor and an instrument of environmental management, IS become indispensable for the management and dissemination of knowledge in order to fulfill the need of organizations for more and better information, thus supporting decision-making and the phases of EMS adoption.

It is necessary to verify how IS can effectively contribute to more effective business environmental management, and which barriers and opportunities exist in this relationship. After that, it is possible to propose potential implications for big data research (Figure 1).

3. Research method
The research was done through a qualitative approach, which is indicated for exploratory research, when there is little material published on the subject in study. In this way, it is appropriate to this research proposal, inherently exploratory, since the integrated study of IS and environmental awareness in organizations is a potential starting point that requires more research (Sarkis et al., 2013).

Regarding the method of this research, we adopted the multiple case study, limited to the study of two cases in companies with the ISO 14001 certification. The adoption of the case study is appropriate for proposed research questions of the type “how” and “why.” This is a way of doing investigative research of current phenomena within its actual context, in situations where the boundaries between the phenomenon and the context are not clearly established (Yin, 2014).

It is believed that the study of two cases is an appropriate strategy for this proposal, as it satisfies the type of research question “how” highlighted by Yin (2014), what is verified, with the established objective of analyzing how the IS can contribute to environmental management, a current phenomenon and with great emergence of the concepts covered. The research organizations were established as two Brazilian companies certified ISO 14001 located in the state of Sao Paulo. Certified companies tend to present an environmental performance higher than the average.

Aligned with the methodology, a case study protocol (Table II) was developed to guide the actions of the researchers and to ensure greater reliability of the results. The process of data collection, in addition to being subsidized by a documentary analysis of the respective environmental policies and in technical information of the organizations, involved also the conduct of interviews based on a guide of questions resulting from the theoretical foundation.
The interviews were directed to professionals responsible for the areas of IS (IT coordinator, project leader) and environmental management (coordinator of the department of environment, environmental control) of each company. The list of questions covered specific points related to the respective area of operation of each of the employees interviewed and, aiming to obtain a contextual analysis interconnected, the guides were supplemented by questions that considered fundamental points for the prospecting of evidence relating to the relationship between IS and environmental management. Table III presents the information on the size and sector of operation of the studied cases, in addition to the dynamics of the data collection.

The data collected were analyzed according to the definition of the study variables: contributions from the IS to the environmental management practices, the opportunities and the emerging challenges of this relationship. The analysis of the data, according to the variables mentioned, can be verified in Table IV.
To reach the ideal number of companies to be part of this case study research, we used the concept of saturation (Eisenhardt, 1989). In this context, a third case study was initiated, however, the results from this third case were too incremental and marginal to be considered further. Additionally, we considered that a qualitative research is

<table>
<thead>
<tr>
<th>Company A</th>
<th>Company B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information system contributions to environmental management</td>
<td>Exhibition of sustainability projects, environmental goals and objectives, risk analysis, environmental indicators and sustainable reports through the sustainability portal</td>
</tr>
<tr>
<td>The information system supports environmental management by conducting procedures such as creating tables, graphs and environmental reports</td>
<td>Control of performance indicators: managers receive data and insert them into the system, these are compiled and generate the sustainability reports</td>
</tr>
<tr>
<td>Storage of documents related to environmental management, as results of environmental analyses, historical and sustainability reports</td>
<td>Sustainability portal supports all documentation of the Environmental Management System (EMS), storing and managing all documentation and manuals</td>
</tr>
<tr>
<td>Assists in the development of spreadsheets, tables and graphs applicable to the area of environmental management</td>
<td>Shared Point supports the approval of suppliers that respect the environment</td>
</tr>
<tr>
<td>Support decision-making on projects in the environmental area, through the information stored</td>
<td>Updates of environmental documents, referred to a chain of approval, and subsequently disclosed in the sustainability portal</td>
</tr>
<tr>
<td>Collaboration in the adoption of ISO 14001 certification through document control</td>
<td>Modules to support waste management and carbon emissions (CO2)</td>
</tr>
<tr>
<td>Management of environmental evidence and dissemination of actions to stakeholders</td>
<td>Management of environmental evidence and dissemination of actions to stakeholders</td>
</tr>
<tr>
<td>Aid in the issuance of sustainability reports and in the control of the validity of documents</td>
<td>Aid in audits by providing relevant information</td>
</tr>
<tr>
<td>Aid in audits by providing relevant information</td>
<td>Internal disclosure of the company’s environmental norms and policies</td>
</tr>
<tr>
<td>External sharing of documents, such as the environmental impact study, assisting in communicating with stakeholders</td>
<td>Exertion of sustainability projects, environmental goals and objectives, risk analysis, environmental indicators and sustainable reports through the sustainability portal</td>
</tr>
<tr>
<td>Aid in sustainability decisions, supporting the creation of new projects, setting goals or environmental objectives according to performance</td>
<td>It performs control of the trainings and qualification of the employees, requirement of ISO 14001, maintaining a history of each employee</td>
</tr>
<tr>
<td>It performs control of the trainings and qualification of the employees, requirement of ISO 14001, maintaining a history of each employee</td>
<td>Inventory management and sustainability reports</td>
</tr>
<tr>
<td>Speed of access to all EMS documentation</td>
<td>Speed of access to all EMS documentation</td>
</tr>
<tr>
<td>Challenges</td>
<td>Opportunity</td>
</tr>
<tr>
<td>Development of the area of information systems, so that environmental aspects are incorporated and the relationship between them is better explored</td>
<td>Consider environmental issues as part of the business as well as information systems, so that the integration between both results in better outcomes for the company in relation to environmental practices</td>
</tr>
</tbody>
</table>

To reach the ideal number of companies to be part of this case study research, we used the concept of saturation (Eisenhardt, 1989). In this context, a third case study was initiated, however, the results from this third case were too incremental and marginal to be considered further. Additionally, we considered that a qualitative research is
recommended for the generation of new insights in a very emerging knowledge field, such as the relation between IS-environmental management with implications for big data.

4. Results

Company A is one of the top five manufacturers of automotive and power generation support batteries in the country. Company B is the largest industrial timber panel company in the southern hemisphere, as well as a Brazilian leader in the market for the production of laminate flooring. Currently, the latter integrates the portfolio of companies present in the Dow Jones Sustainability Index.

Both organizations consider the sustainable aspect in their missions and values, demonstrating that the environmental dimension is a function of great importance in their business, a fact confirmed by the environmental practices applied to the management of each of them, as shown in the following list.

Some environmental practices of the companies studied:

(1) Company A:
- environmental policy defined and aligned with the environmental commitment of the organization, which integrates the business mission;
- environmental management supported by EMS ISO 14001 certificated;
- reverse logistics practice;
- air pollution control plan;
- air quality monitoring plan;
- groundwater monitoring plan;
- soil and vegetation monitoring plan;
- consumption of electrical energy generated by low environmental impact sources;
- storage and correct disposal of waste;
- environmental awareness of employees through specific training; and
- documentation and analysis of environmental indicators.

(2) Company B:
- environmental policy defined with the reaffirmation of the commitment as an environmentally responsible company;
- environmental management supported by EMS ISO 14001 certificated;
- rational and sustainable use of natural resources, raw materials and inputs necessary for production processes;
- development and supply of products that enable the rational use of natural resources;
- prevention of pollution and environmental risks arising from production, including operations focused on reducing greenhouse gas emissions;
- compliance with the law applicable to activities, products and services concerning the company;
- establishment of procedures ensuring the non-use of inputs of illegal origin;
- protection of biodiversity, springs and water courses, as well as conservation of cultivated soil, measures inherent in the management of forestry plantations;
- waste management converging on the concepts of reduction, recycling and reuse;
- implementation of training and qualification programs that lead to the adoption of behaviors of respect to the environment;
- maintenance of communication channels with stakeholders on environmental, social, products and services aspects; and
- documentation and dissemination of scopes and results achieved in the attendance of social and environmental commitments, voluntarily assumed by the company.

The company A has an integrated information system which, according to the respondent, “carries out customizations pertinent to the working environment, covering much of the processes and sectors,” but still does not contemplate all areas, such as the environmental. Therefore, it has some parallel controls and auxiliary systems. Although the system does not integrate the environmental aspect, there is support for environmental management, implemented through the storage and control of documents, as all documentation pertinent to the EMS, results of analyses and environmental historical and assist in creating sustainability reports. The information technology area of the company helps with the development of spreadsheets, tables and graphs applicable to the area of environmental management, as needed. As for the dissemination of environmental practices, it is carried out internally through informative panels and externally through the corporate website. There are long-term projects for the implementation of an information system that contemplates the environmental aspect, spanning from the control of its environmental protection practices to the dissemination of them.

It was found that company B is in a more advanced stage regarding to IS due to the availability of a wider range of functionalities, which covers organizational operations as a whole, including the control and application of environmental management practices. There is an ERP system that contemplates the environmental aspects. There are modules dedicated especially to environmental management, such as environment, health and safety, document management service and environmental compliance. In addition, there is an internal network with a portal of sustainability, which exposes indicators, environmental goals and sustainability reports.

The analysis of the data allows to note that a more advanced stage in relation to the IS is a factor that implies a more effective contribution of the IS in the environmental management practices. Thus, the existence of a more robust information system such as an ERP, which integrates all the data and processes of an organization and incorporates the most diverse departments, determines a cross-collaboration of IS for environmental practices.

In company B, this collaboration begins with the support and storage of the documentation involved in the EMS, includes control and management of waste and carbon emissions, and also incorporates the environmental evidence, which is related to sustainability reports and dissemination of environmental actions to stakeholders. On the other hand, there are still procedures carried out in parallel with the main system, such as the management of non-conformity and the creation of graphs. Still on the company B, it was found that one of the challenges in the relationship between IS and EM is the integration of the employees with the systems, which incorporate the most diverse aspects and processes of the company. The resistance, or even the lack of familiarity of the employees with the information system can be a factor of direct interference to the contribution of the IS to the different areas of the company, including for the environmental management, considering that the supply of data and information is critical in the search for a systemic control of the productive processes and environmental practices.

Finally, the opportunities identified in both companies are mainly related to the consideration of environmental issues as part of the business, the same way that the IS are,
so that the integration between the two of them implies better results for the company as regards environmental practices. Table IV systematizes the main contributions of the IS to environmental management in the cases studied, as well as the challenges and opportunities of this relationship.

5. Discussions

5.1 Managerial and academic implications

The results showed that the alignment between the IS and the environmental aspect generates several contributions to the environmental management of companies. Thus, as suggested by Dao et al. (2011) and Ryoo and Koo (2013), it was verified that this alignment contributes to the improvement of the environmental capacity of the company, while ensuring that IS help reaching environmental goals.

Likewise, it was confirmed that ERP systems and web portals, which incorporate environmental aspects, support the companies in the collection, dissemination, promotion and communication of their environmental practices to stakeholders (Prajogo et al., 2014; Dao et al., 2011; Chen et al., 2008).

It was also verified that the more advanced the information system adopted by the company is, resulting in a greater integration of the different activities and processes related to the organization, the more effective the IS contribution to the environmental management practices and, consequently, better the management of environmental aspects by the organization. From this observation, Figure 2 was elaborated.

The data analysis made possible to describe the contributions of IS to environmental management in the cases studied, which allowed them to be related to the three categories of environmental practices distinguished by González-Benito and González-Benito (2006). The support of IS occurs in all practices categorized by the authors. Based on the research results and the practices of González-Benito and González-Benito (2006), it is possible to propose a framework to understand the relationship between them (Figure 3).

As could be seen, IS provide support for organizational practices by enhancing the decision-making process of creating environmental projects and in the training/qualification of employees. Considering the big data revolution, it implies that information management capability offers a high potential of boosting the development of environmentally sustainable organizational initiatives and business by exploiting big data applications and resources (Jeble et al., 2018).

Regarding operational practices, IS assist the selection and supervision of suppliers based on green criteria, and in the management and control of environmental indicators in production processes. Yang et al. (2018) highlight the importance of integrating IS and supply chains in the deployment of green considerations in both areas. As the number of partners of the organizations increases and the amount of data about environmental impacts of their supply chain also raises, big data may have the ability to evaluate and organize the environmental information originated by all the supply chain partners (Dubey et al., 2017).
At last, it is argued that effective communication is a key factor to implement sustainable practices in an organization successfully (Seuring and Müller, 2008; González-Benito and González-Benito, 2006). The findings show that IS help in the communication and dissemination of environmental compliance actions to internal and external stakeholders. In this sense, with the growing number of communication channels and transparency that the new technologies are providing, researchers will need to analyze the influence that big data may have on the implementation of communicational environmental practices.

5.2 Implications for big data research: a research agenda

Based on the above discussion and the positive linkage between IS and environmental management, it is possible to highlight that big data would have a number of implications for the IS-environmental management interaction. This allow us to propose a research agenda. The main questions that emerge from IS-environmental management for big data research are:

- How can companies use big data analytics in order to improve the relation between IS-environmental management?
- How will big data affect the adoption of practices concerning organizational environmental practices?
- How will big data affect the adoption of operational environmental practices?
• How will big data influence the implementation of communicational environmental practices?
• Are big data-capable companies more likely to have a better environmental management?
• Are IS applied to environmental management ready for the big data revolution?
• What would be the challenges of applying big data for environmental management?
• What would be the competitive advantages from big data for environmental management strategies?
• How to use big data for ISO 14001 implementation?
• How to use big data for ISO 14001 maintenance and continuous improvement?

These and other questions can be useful for further understanding the relation between IS and environmental management in the twenty-first century.

6. Conclusion
This research aimed to identify the contributions of IS to environmental management practices in the cases analyzed, as well as the opportunities and challenges with implications for big data research. Starting from a study of multiple cases with two companies, it was possible to reach the objectives initially proposed. The most relevant result of this research is the identification of how IS have effectively contributed to environmental management. Indeed, we believe that big data could boost this already positive relation between IS-environmental management.

Although the current literature highlights the importance of environmental management and IS in organizations, there is little in the state-of-the-art about the relationship between both (Fiorini and Jabbour, 2017; Ryoo and Koo, 2013; Dao et al., 2011; Sarkis et al., 2013; Melville, 2010). Thus, this research contributes specifically to the production of evidence of this relationship, highlighting how the IS support the environmental management of the company, still with the present limitations in this study.

In addition, another contribution is the finding that companies at a more advanced stage in IS, using integrated management systems such as ERPs aligned with the environmental aspect, receive a greater contribution of the IS in the environmental management practices, being able to include all the categories established by González-Benito and González-Benito (2006). The framework on the relationship between environmental practices and IS, presented in Section 5, represents an advance in the literature on this subject.

At last, as a final contribution we presented a research agenda on the implications of big data for IS-environmental management relation. This research agenda can be useful for future researchers in big data and sustainability. Accordingly to Dubey et al. (2017) the use of big data for promoting environmental sustainability is at the forefront of the management research agenda.

The results of this research may be useful to business leaders and researchers in the areas of IS and environmental management. For business leaders, the importance of considering IS as tools to assist environmental management is highlighted, in line with the search for improvement in the capacity and environmental performance of their companies. To the researchers of the areas of IS and environmental management, the present study adds findings and evidence to a subject still not very explored. It is expected that the results obtained could foster new initiatives in organizations for better environmental performance, by narrowing the relationship between environmental management and IS and exploring the implications for future big data applications.
As a limitation of the research, it is highlighted that the results cannot be generalized to a large extent, since this is based on only two cases, referring to companies located in Brazil. From this study, it is possible to suggest new investigations in the area. A survey with a greater number of cases can be considered, as well as a deeper investigation into how IS are useful in each of the organizational, operational or communicational practices. Issues such as the use of IS and big data technologies in increasing sustainability and the correlation between informational capacity and environmental management in the organizations also require future research. Finally, considering the limitation that our study is only based on Brazil, it is suggested to develop further research on the role of IS-environmental management in multinational companies (Kumar et al., 2018; Singh and Delios, 2017) and in emerging economies, such as China (Gaur et al., 2018) and India, in order to enhance the body of knowledge on the topic.

References


About the authors
Paula de Camargo Fiorini is Researcher at UNESP-Sao Paulo State University. Her research interests are sustainable operations, sustainable information systems, and information technology.

Charbel Jose Chiappetta Jabbour is Associate Professor at Montpellier Business School. His research interests are green management, the “soft side” of sustainability, and green human resources. He has been serving as Associate Editor for the Journal of Cleaner Production. Charbel Jose Chiappetta Jabbour is the corresponding author and can be contacted at: c.chiappetta-jabbour@montpellier-bs.com

Ana Beatriz Lopes de Sousa Jabbour is Associate Professor at Montpellier Business School. Her research interests are green management, industry 4.0 and circular economy. She has been serving as Editor for the Latin American Journal of Management for Sustainable Development.

Nelson Oliveira Stefanelli is early Career Researcher at University of Sao Paulo, Ribeirao Preto, Brazil. His research interests are green management, industry 4.0 and circular economy.

Yudi Fernando is Associate Professor at Faculty of Industrial Management, Universiti Malaysia Pahang, Malaysia, and at Management Department, BINUS Online Learning, Bina Nusantara University, Indonesia. He has published on sustainable operations.

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Transforming big data into knowledge: the role of knowledge management practice

Roberto Chierici and Alice Mazzucchelli
Department of Business and Law, Universita degli Studi di Milano-Bicocca, Milano, Italy
Alexeis Garcia-Perez
Coventry University, Coventry, UK, and
Demetris Vrontis
School of Business, University of Nicosia, Nicosia, Cyprus

Abstract

Purpose – The purpose of this paper is to empirically investigate how big data collected from social media contribute to knowledge management practices, innovation processes and business performance.

Design/methodology/approach – The study used 418 questionnaires collected from firms that actively invest in marketing, advertising and communication in the Italian market. The hypotheses testing and analysis were conducted using structural equation modeling.

Findings – The results reveal that customers’ data gathered from social media produce different effects on knowledge management practices and firms’ innovation capacity. Furthermore, increased innovation capacity turned out to affect customer relationship performance directly, while it contributes to gain better financial performance only when it is used to gain relational outcomes.

Originality/value – The outcomes of the study help firms to develop a clear understanding about which big data retrieved from social media can be useful to improve their knowledge management practices and enhance their innovation capacity. Moreover, by investigating the mediating role of big data knowledge management in the context of social media knowledge acquisition and innovation capacity, this study also extends the mediation variables used to understand the relationship between knowledge capabilities and practices and innovation constructs.

Keywords Firm performance, Market orientation, Social media, Innovation capacity, Big data knowledge management, Customer collaboration

Paper type Research paper

1. Introduction

Introducing new customer-centric tools, social media have transformed the way firms communicate and interact with customers. Posts, likes, tweets, digital pictures and videos, geotags are only some sources of big data that firms are collecting, storing, managing and analyzing to understand how they can serve customers better (Fosso Wamba et al., 2015, 2017; Khan and Vorley, 2017; Pauleen and Wang, 2017). In 2017, more than 3bn people worldwide, and 34m people in Italy, actively used social media each month (Kemp, 2018), generating a huge amount of data that can represent an endless and continuously updated source of information. Nowadays, social media represent an external source of knowledge thanks to which firms can assume data-driven decisions, improving their innovation capacity and staying ahead of competition (Bean and Kiron, 2013; Mukherjee et al., 2017; Nuruzzaman, Gaur and Sambharya, 2018). By managing big data, firms can derive information useful to enhance their operational efficiency, innovate their products/services and processes, reinforce their relationships with customers and, consequently, enhance their overall performance (Fosso Wamba et al., 2017). According to the resource-based view and the knowledge management literature, being market oriented and actively collaborate with customers allow firms to develop intangible assets, such as
knowledge and market sensitivity, that can be deployed to innovate and increase firms’ performance (Gaur et al., 2011). However, extracting knowledge from big data, integrating them within firms’ processes, and turning insights into decision-making actions poses significant challenges (Chen et al., 2012; Contractor et al., 2016; Nuruzzaman, Gaur and Sambharya, 2018) and firms are struggling on understanding how they can effectively exploit all these information to achieve higher level of innovation capacity and, consequently, improve their performance.

For this reason, both academics and practitioners have deeply investigated big data and social media in order to evaluate how these phenomena are changing the dynamics of the competitive environment (Erevelles et al., 2016; Fosso Wamba et al., 2015; Rothberg and Erickson, 2017). However, how big data gathered from social media can contribute to knowledge management practices, innovation processes and business performance remains largely unexplored. To bridge this gap, the study proposes a conceptual model that aims to analyze the causal relationships among social media market orientation, in terms of both proactive and reactive orientation, social media customer collaboration, big data knowledge management, innovation capacity and firms’ performance.

In this perspective, the contributions of the research are the following. First, by analyzing both the direct and indirect effects, the study demonstrates that different ways to acquire customer-related data from social media differently affect knowledge management practices and firms’ innovation capacity. Second, the research examines the mediating role of big data knowledge management in the context of social media knowledge acquisition and innovation capacity, by extending the mediation variables used in previous studies to understand the relationship between knowledge capabilities and practices and innovation constructs. Finally, the study tests the causal relationship between innovation capacity and business performance, obtaining mixed results. If, on the one hand, innovation capacity directly influences customer relationship performance, on the other, innovation capacity seems to affect financial performance only when it is used to gain relational outcomes.

The paper is structured as follows. In the next section, the study provides the theoretical background and develops the research hypotheses. Then, the methodology used and the results obtained are presented. Following this, the theoretical and practical implications of the study’s findings are discussed and limitations and directions for future research are presented.

2. Literature review
2.1 Market orientation, big data knowledge management and innovation capacity
Market orientation can be intended as the process firms adopt to systematically generate and disseminate customers’ data and intelligence in order to understand current and future customers’ needs (Kohli and Jaworski, 1990), and it is related to firms’ attitude to rely on information about customers to define market strategies and create superior value (Narver and Slater, 1990). More in detail, market orientation has been conceptualized on the basis of two different approaches, the behavioral approach and the cultural approach (Gaur et al., 2011). According to the behavioral approach, market orientation is a set of ongoing activities that contributes to enhance customer relationship performance, including knowledge generation and dissemination, and firms’ ability to promptly respond to customers’ instances (Kohli and Jaworski, 1990). Instead, the cultural approach posits that market orientation consists of three different components (customer orientation, competitor orientation and interfunctional coordination) and two decision criteria (long-term focus and profitability) that let firms to create a superior value for their customers continuously and gain a competitive advantage (Narver and Slater, 1990).
In a dynamic-capabilities perspective (Eisenhardt and Martin, 2000), market orientation allows firms to develop a deeper understanding of customers’ wants and needs (Hult and Ketchen, 2001; Jaworski and Kohli, 1993), supports firms in selecting the most effective resource combinations to meet market conditions (e.g. Slater and Narver, 1998) and, consequently, it can be a source of superior competitive advantage.

Narver et al. (2004) suggest that market orientation can be reactive or proactive. Firms that implement a reactive market orientation try to identify, understand and satisfy the expressed needs of customers, while those firms that adopt a proactive market orientation are more focused on recognizing and responding to customers’ latent needs.

While some studies have suggested that to find out new market opportunities and exploit them, firms have to adopt at least one of the two market orientation main approaches (Marvel and Lumpkin, 2007; Nguyen et al., 2015), others convey that firms should practice both proactive and reactive market orientation to acquire data about customers and used them to empower their knowledge (Kristensson et al., 2008; Nguyen et al., 2015; Ordanini and Maglio, 2009). In the new digital domain, being customer oriented is a crucial competence for firms and social media are a primary source of big data that firms can adopt to understand customers’ expressed wants and latent needs, collaborate with them, co-create products and services that meet their exigencies (Gaur, 2006; Gupta et al., 2010; Stefanou et al., 2003). However, firms’ ability to transform big data acquired from social media into knowledge depends on the extent to which they are able to evaluate new information and opportunities, use them to improve their knowledge capabilities, recombine existing information, generate new solution and add value through knowledge management practices (Cambra-Fierro et al., 2011; De Dreu and West, 2001; Nguyen et al., 2015; Tiwana and McLean, 2005). In this perspective, even if proactive and reactive market orientations should be simultaneously adopted to acquire and exploit information, if and how these two approaches contribute to generate customer-related knowledge remains under-investigated (Nguyen et al., 2015; Ozkaya et al., 2015). Therefore, based on the above discussion, it is expected to find a positive effect of both proactive social media market orientation and reactive social media market orientation on big data knowledge management:

H1. Proactive social media market orientation has a positive influence on big data knowledge management.

H2. Reactive social media market orientation has a positive influence on big data knowledge management.

Being oriented toward markets sustains firms in developing new ideas (Hurley and Hult, 1998). Market-oriented activities, if combined with appropriate capabilities, may contribute to acquire advantages in product and process innovations (Slater and Narver, 1998). Furthermore, thanks to the information acquired through social media, firms can improve their innovation capacity and better address customers’ needs by developing new ideas or products (Slater and Narver, 1998). Despite previous studies have focused on market orientation, both proactive and reactive, and innovation (Narver et al., 2004), they have found mixed results concerning the effect of market orientation on innovation (Nguyen et al., 2015). More in detail, some authors have found a positive impact (Atuahene-Gima, 2005; Gotteland and Boule, 2006; Kam Sing Wong and Tong, 2012; Narver et al., 2004; Vega-Vázquez et al., 2012), others have demonstrated a negative effect (Frambach et al., 2003; Perry and Shao, 2005) or no effect (Im and Workman, 2004; De Luca et al., 2010). Moreover, literature has suggested that reactive market orientation can contribute to the development of incremental innovations, while proactive market orientation, leading to deeper insights into customers’ needs, can be exploited by firms to develop radical innovations (Deshpandé et al., 1993; Narver et al., 2004). Extant literature suggests that proactive orientation allows firms to disrupt their existing capabilities and create new ones that could be exploited to develop radical innovation and carry out new
products/services or processes, while reactive orientation allows firms only to enhance their existing capabilities and use them to develop incremental innovations (Forsman, 2011; Nuruzzaman, Singh and Pattnaik, 2018).

Being market oriented lets firms respond to changes, boosting their innovation capacity through continuous innovation (Deshpandé et al., 1993; Fidel et al., 2016). Therefore, it is expected to find a positive relationship between both forms of market orientation and innovation capacity:

H1a. Proactive social media market orientation has a positive influence on innovation capacity.

H2a. Reactive social media market orientation has a positive influence on innovation capacity.

2.2 Customer collaboration, big data knowledge management and innovation capacity

Customer collaboration via social media refers to information gained from customers that actively interact and collaborate with firms in the value co-creation process (Constantinides et al., 2009). In value co-creation, customer is a fundamental player, performing as an active co-inventor of value (Lusch et al., 2007; Vargo and Lusch, 2008; Vega-Vazquez et al., 2013), and social media, allowing interactions and the sharing of information, interests and opinions between customers and firms and among peers, have facilitated the value co-creation process (Harrison and Barthel, 2009). Social media environment has enabled firms to directly and continuously collaborate with customers and develop a learning process from them (Sawhney et al., 2005). Consequently, by exploiting social media, firms can shape relationships with existing customers, acquire new customers, and set up communities that interactively collaborate to identify and understand existing and latent needs and develop solutions for customers (Sashi, 2012). Collaborating with customers through social media represents a primary determinant for firms to acquire customer-related data that, in turn, require to be adequately managed in order to gain a customers’ knowledge useful to support co-creation processes (Bharati et al., 2014; Fidel et al., 2016). In a dynamic-capability perspective, collaborating with customers lets firms acquire that external knowledge required to generate new learning and accumulate experience useful to develop an enduring source of competitive advantage (Alegre et al., 2011; Marsh and Stock, 2006).

Thus, the study investigates the relationship between customer collaboration through social media and big data knowledge management practices, and posits that:

H3. Social media customer collaboration has a positive influence on big data knowledge management.

Previous studies have investigated customer collaboration within the innovation process. According to Wind and Mahajan (1997), firms that cultivate strong collaborations with their customers acquire useful information that can be exploited for the development of successful innovations. Customers no longer play a passive role, merely answering questions or allowing observations, but actively take part to the innovation process as valuable co-creators. In other terms, by actively collaborate with customers, firms have the opportunity to deploy them as a strategic resource that can be involved to jointly discover customers’ latent needs and, consequently, empower firms’ innovation capacity (Ordanini and Parasuraman, 2011; Vargo and Lusch, 2008). As social media allow firms to establish collaborative conversations and enhance relationships with customers (Greenberg, 2010; Trainor, 2012), social media represent a tool through which cooperate with customer and support the value co-creation process (Trainor et al., 2014).
In this perspective, since numerous research works report a positive relationship between customer collaboration and innovation capacity, the study posits:

H3a. Social media customer collaboration has a positive influence on innovation capacity.

2.3 Big data knowledge management and innovation capacity

Knowledge management has been identified as an important antecedent of innovation (Carneiro, 2000; Dove, 1999), even if its effect on innovation is hard to determine (Darroch and McNaughton, 2002). However, previous studies convey that knowledge generation and dissemination play a crucial role in gaining a sustainable competitive advantage, such as innovation, because of their uniqueness to the firm (Day, 1994; Grant, 1996). In fact, according to Meso and Smith (2000), knowledge management is “the process of capturing the collective expertise and intelligence in an organization and using them to foster innovation through continued organizational learning” (p. 225). Knowledge management grasps the changes occurring in the environment and supports firms in integrating, building and reconfiguring their competences. In this perspective, knowledge management has been associated with firms’ practices like organizing knowledge repositories, adopting technologies that allow collecting data from internal and external sources, and developing mechanisms to share and transfer knowledge (Darroch and McNaughton, 2002; Gupta et al., 2000). In recent years, the knowledge process and practices have undergone a revolution since Web 2.0 and social media have altered the way through which firms create, share and capture data and, at the same time, have allowed firms to access big data that, if adequately managed, become an additional valuable knowledge asset (Erickson and Rothenberg, 2014; von Krogh, 2012). Knowledge represents a basis for the development of a competitive advantage (Lusch et al., 2007) and, contributing to the enhancement of firms’ innovation capacity, it results as a key element of firm competitiveness (De Clercq and Arenius, 2006; Nonaka, 1994). Moving from the assumption that innovation is the application of knowledge (Fidel et al., 2016), and that these two concepts are strictly connected one to each other, this study posits:

H4. Big data knowledge management has a positive influence on innovation capacity.

2.4 Innovation capacity and firm performance

Innovation capacity can be defined as the firms’ ability to develop and realize new processes and value propositions (Hurley and Hult, 1998) that satisfy customers’ current and latent needs (Adler and Shenbar, 1990). According to Deshpandé et al. (1993), innovation capacity is a source of competitive advantage because it allows firms to adapt themselves to the dynamic environment wherein they operate and compete. Developing and exploiting innovation capacity is not only a strategic choice but it is also a crucial aspect of firms’ long-term competitiveness (Singh and Gaur, 2013).

Even if innovation is a high-risk and resource-consum ing activity that requires significant R&D investments and a specific allocation of managerial and financial resources and, in the short-term, it could lead to performance not as much positive as expected (Lee et al., 2017), previous studies have demonstrated the importance of innovation in contributing to firms’ long-run competitiveness and the existence of a positive relationship between innovation constructs and the different dimensions of firms’ performance (Calantone et al., 2002; Hitt et al., 1997). Furthermore, Hurley and Hult (1998) reveal that firms’ capacity to innovate, covering different strategic areas and business units, from product design to marketing, affects firms’ competitiveness. Moreover, a number of studies note that generating and utilizing knowledge to improve firms’ innovation capacity leads firms to achieve higher performance (Ozkaya et al., 2015; Palacios-Marqués et al., 2015).
In fact, the capacity to better respond to customers’ needs through the development of innovative products and services enhances both relational outcomes, such as customer satisfaction, customer loyalty and customer retention, and financial outcomes, such as firms’ sales, profitability and market share (Fidel et al., 2015, 2016; Kostopoulos et al., 2011).

Thus, it can be assumed that innovation capacity is essential for firms to achieve superior business performance outcomes, in terms of both customer relationship performance and financial performance. Therefore:

- **H5.** Innovation capacity has a positive influence on customer relationship performance.
- **H6.** Innovation capacity has a positive influence on financial performance.

### 2.5 Customer relationship performance and financial performance

Customers play a crucial role for firms to compete and succeed in the actual scenario. Since projecting and designing new products or services in collaboration with customers lead firms to propose an offer that is more highly valued by customers (Kristensson et al., 2008), firms are trying to involve them in co-creating new products. This effect influences also firms’ financial performance because customers who have access to products and services that respond to their needs and exigencies tend to establish a relationship with the brands and, consequently, to generate more purchases over times (Reinartz et al., 2004). Consequently, it emerges that customers have different economic value to firms that, in turn, are interested in implementing tools, technologies and processes that can be used to establish better and longer relationship with customers (Zablah et al., 2004). In this perspective, firms have to understand how they are performing with their customers. Gupta and Zeithaml (2006) suggest that customer metrics can be classified as stated preferences, which are those unobserved preferences such as customer satisfaction, and revealed preferences, which are those related to customers’ behavior such as customer loyalty and retention. Both these categories refer to customers’ attitude to get engaged with firms and, thus, it may be expected they affect firms’ profitability (Verhoef et al., 2010). Therefore:

- **H7.** Customer relationship performance has a positive influence on financial performance.

### 2.6 The mediation effects of big data knowledge management

Previous arguments provide the theoretical foundations for the final hypotheses of the study that assume big data knowledge management acts as a mediation variable of the relationships between proactive social media market orientation, reactive social media market orientation, social media customer collaboration and innovation capacity. Previous literature suggests that acquiring data and transforming them into knowledge contributes to strength firms’ innovation capacity (Taghizadeh et al., 2018). For instance, some authors have investigated the relationship between knowledge acquisition and innovation by considering knowledge competence (Ozkaya et al., 2015), instrumental use of information (Gotteland and Boulé, 2006), organizational learning (Zhou et al., 2005), and research and development effectiveness (De Luca et al., 2010) as mediation variables. All these studies have considered different aspects of knowledge management to explain the complex effect of knowledge acquisition orientation on innovation constructs. In this perspective, the study assumes that big data gathered through social media market orientation, in terms of both proactive and reactive, and social media customer collaboration are transformed into knowledge through big data knowledge management, and this, in turn, improve firms’ innovation capacity. Thus:

- **H8.** Big data knowledge management mediates the positive relationship between proactive social media market orientation and innovation capacity.
H9. Big data knowledge management mediates the positive relationship between reactive social media market orientation and innovation capacity.

H10. Big data knowledge management mediates the positive relationship between social media customer collaboration and innovation capacity.

3. Method

3.1 Sampling and data collection

The objective of this research is to investigate the role and the impact of social media market orientation, both proactive and reactive, and social media customer collaboration on big data knowledge management, innovation capacity and firm performances, as outcomes. Data on such constructs were collected through a self-administered web-based questionnaire dispatched to managers of firms that operate in Italy and use at least one social network to communicate and interact with their customers.

The chosen respondents for the questionnaire were managers whose holistic view enables them to provide reliable responses about their organizations’ activities (Hambrick and Mason, 1984). In addition, the research was carried out within the Italian context because, according to We Are Social 2018 report, both firms and consumers daily use social media to share information, experiences and engage with brands (Kemp, 2018) and, consequently, the Italian market represents a suitable context for social media research.

The questionnaire, in which respondents self-reported their answers, was developed and divided into two sections. The first section was dedicated to study the seven constructs adapted from previous literature and revised to fulfill the research aim; the second part addressed the characteristics of the investigated firms. Prior to the data collection, a pre-test was conducted with ten academics and managers to check the contents of the questionnaire and the appropriateness of the questions.

In order to achieve a large number of managers from a wide range of industries and different business sizes, the invitation to fill in the questionnaire was sent to 1,565 managers’ e-mail contacts sourced through a collaboration with LeFAC, a database that collects information and insights about firms that actively invest in marketing, advertising and communication in the Italian market. From June to September 2016, 418 questionnaires were returned in a completed form, which represents a response rate of 26.7 percent. This response rate is in line with the common standards for web-based questionnaires administered to firms’ managers (Anseel et al., 2010; Cycyota and Harrison, 2006). Since data collection was performed through an online questionnaire, respondents were not allowed to move forward to the following question if they did not answer to the previous one. Hence, the study did not provide any missing value.

3.2 Measures

All the measurement scales for operationalizing each construct of the conceptual model have been previously validated. The study uses a seven-point Likert scale to measure all the constructs’ items.

Based on the research of Jaworski et al. (2000), Narver et al. (2004), Ordanini and Maglio (2009) and Nguyen et al. (2015), the study measures proactive social media market orientation and reactive social media market orientation using a four-item scale and a three item-scale, respectively. Proactive social media market orientation items measure firms’ ability in using social media to discover customers’ latent needs, exploit new market opportunities, and cannibalize existing offerings. Reactive social media market orientation items examine firms’ ability in using social media to acquire and generate information regarding existing customers’ needs, exigencies and satisfaction.
Further, adapting the scale proposed by Fidel et al. (2015, 2016) and Santos-Vijande and Álvarez-González (2007), the study investigates social media customer collaboration and innovation capacity using a four-item scale each. Social media customer collaboration explains firms’ ability to acquire knowledge from social media through the continuous interaction and conversation with their customer, while innovation capacity is related to firms’ ability to develop new ideas or products using information derived from big data management.

Drawing from dynamic capability theory (Barney, 1991; Nielsen, 2006; Teece, 2009; Teece et al., 1997) and Alegre et al. (2011), Fidel et al. (2015, 2016) and O’Connor and Kelly (2017) works, the study measures big data knowledge management using a seven-item scale. Big data knowledge management describes firms’ abilities and capabilities to exploit big data-enabled technologies and infrastructures to gain and share a deeper knowledge about customers.

Finally, the study uses a five-item scale to measure customer relationship performance (Rapp et al., 2010; Trainor et al., 2014) and a three-item scale to evaluate financial performance (Grissemann et al., 2013; Ozkaya et al., 2015). Comparing firms’ performance to competitors, customer relationship performance assesses firms’ success in satisfying and retaining customers, while financial performance construct evaluates firms’ sales growth, profitability and market share.

The Appendix presents the scale items of each construct analyzed in this study.

3.3 Data analysis
Using LISREL 8.80, the structural equation modeling (SEM) technique was applied in order to empirically test the relationship of proactive social media market orientation, reactive social media market orientation, social media customer collaboration, big data knowledge management, innovation capacity, customer relationship performance and financial performance.

4. Results and hypotheses testing
4.1 Measurement model
Using SPSS and LISREL 8.80, the study estimates Cronbach’s alphas (CA), item-to-total correlations, and confirmatory factor analysis to test reliability, convergent validity and discriminant validity of each construct (Anderson and Gerbing, 1988; Jöreskog and Sörbom, 2005).

All Cronbach’s α values exceed the suggested threshold of 0.70 (Bagozzi and Yi, 1988; Nunnally and Bernstein, 1994), ranging from 0.787 to 0.915, signifying an acceptable reliability of each of the study constructs.

With regard to convergent validity test, all item loadings are greater than the recommended threshold of 0.50 (Hair et al., 2006), all the composite reliability (CR) values are higher than the minimum threshold of 0.70 (Bagozzi and Yi, 1988; Nunnally and Bernstein, 1994), and all the average variance extracted (AVE) values exceed the recommended threshold of 0.50 (Fornell and Larcker, 1981). Together, these results indicate an adequate convergent validity for all constructs. Furthermore, all AVE values are greater than the squared correlations of the constructs, showing a good level of discriminant validity of the measurement scales (Fornell and Larcker, 1981). Thus, also the discriminant validity of the constructs is supported.

Table I shows reliability, convergent and discriminant validity examinations, and Table II presents the correlation matrix.

Finally, due to the use of a structured questionnaire in which respondents self-reported their answers, several approaches to minimize the potential for common biased effect are used. In particular, by pre-testing the survey, the item statements were clarified to reduce items ambiguity and the items related to the dependent variables were not located near to the independent ones. Moreover, as suggested by Podsakoff et al. (2003), the Harman’s
A single-factor test was carried out. All measurement items were loaded into an exploratory factor analysis, using principal components extraction and unrotated factor solution, to check if the variance of all items was explained by only one component. No evidence of common method bias was found.

### 4.2 Structural model

The structural model results, including the relationships among constructs, overall explanatory power, completely standardized coefficients and t-values are presented in Table III. The structural model has an acceptable fit with the empirical data, with $\chi^2$ 1,498.25127; degrees of freedom 392; $\chi^2$/df 3.822; root mean square error of approximation 0.089525; comparative fit index 0.96413; standardized RMR (SMRM) 0.14341. All items load significantly on their assigned latent constructs.

#### Table I.

CFA – Reliability, convergent and discriminant validity examinations

<table>
<thead>
<tr>
<th>Constructs</th>
<th>$\alpha$</th>
<th>CR</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proactive social media market orientation</td>
<td>0.859</td>
<td>0.875</td>
<td>0.639</td>
</tr>
<tr>
<td>Reactive social media market orientation</td>
<td>0.856</td>
<td>0.857</td>
<td>0.668</td>
</tr>
<tr>
<td>Social media customer collaboration</td>
<td>0.882</td>
<td>0.885</td>
<td>0.663</td>
</tr>
<tr>
<td>Big data knowledge management</td>
<td>0.898</td>
<td>0.901</td>
<td>0.566</td>
</tr>
<tr>
<td>Innovation capacity</td>
<td>0.787</td>
<td>0.801</td>
<td>0.512</td>
</tr>
<tr>
<td>Customer relationship performance</td>
<td>0.915</td>
<td>0.919</td>
<td>0.696</td>
</tr>
<tr>
<td>Financial performance</td>
<td>0.906</td>
<td>0.907</td>
<td>0.764</td>
</tr>
</tbody>
</table>

#### Table II.

Correlation matrix

<table>
<thead>
<tr>
<th>Constructs</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Proactive social media market orientation</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) Reactive social media market orientation</td>
<td>0.373</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Social media customer collaboration</td>
<td>0.574</td>
<td>0.348</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Big data knowledge management</td>
<td>0.624</td>
<td>0.265</td>
<td>0.525</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) Innovation capacity</td>
<td>0.571</td>
<td>0.654</td>
<td>0.493</td>
<td>0.488</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) Customer relationship performance</td>
<td>0.351</td>
<td>0.464</td>
<td>0.303</td>
<td>0.300</td>
<td>0.615</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>(7) Financial performance</td>
<td>0.279</td>
<td>0.368</td>
<td>0.240</td>
<td>0.238</td>
<td>0.488</td>
<td>0.662</td>
<td>1.000</td>
</tr>
</tbody>
</table>

#### Table III.

Structural relationships and hypotheses testing

<table>
<thead>
<tr>
<th>Path</th>
<th>Completely std $\beta$ and $\gamma$ t-value</th>
<th>Hypotheses test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proactive SM market orientation $\rightarrow$ big data knowledge management</td>
<td>0.48047 7.60512***</td>
<td>Supported</td>
</tr>
<tr>
<td>Proactive SM market orientation $\rightarrow$ innovation capacity</td>
<td>0.19626 3.26636***</td>
<td>Supported</td>
</tr>
<tr>
<td>Reactive SM market orientation $\rightarrow$ big data knowledge management</td>
<td>$-0.00093$ $-0.01916$</td>
<td>Not supported</td>
</tr>
<tr>
<td>Reactive SM market orientation $\rightarrow$ innovation capacity</td>
<td>0.60888 10.97465***</td>
<td>Supported</td>
</tr>
<tr>
<td>SM customer collaboration $\rightarrow$ big data knowledge management</td>
<td>0.25015 4.39176***</td>
<td>Supported</td>
</tr>
<tr>
<td>SM customer collaboration $\rightarrow$ innovation capacity</td>
<td>0.08382 1.59638</td>
<td>Not supported</td>
</tr>
<tr>
<td>Big data knowledge management $\rightarrow$ innovation capacity</td>
<td>0.16058 2.89399***</td>
<td>Supported</td>
</tr>
<tr>
<td>Innovation capacity $\rightarrow$ customer relationship performance</td>
<td>0.61503 10.78117***</td>
<td>Supported</td>
</tr>
<tr>
<td>Innovation capacity $\rightarrow$ financial performance</td>
<td>$-0.06812$ $-1.46505$</td>
<td>Not supported</td>
</tr>
<tr>
<td>Customer relationship performance $\rightarrow$ financial performance</td>
<td>0.90387 17.05798***</td>
<td>Supported</td>
</tr>
</tbody>
</table>

**Note:** ***$p < 0.01$**
$t$-values indicate that seven out of ten research hypotheses presented in Figure 1 are supported. The relationships between proactive social media market orientation and big data knowledge management and between proactive social media market orientation and innovation capacity are positive and significant ($\gamma = 0.48047$, $t = 7.60512$, $p < 0.01$; $\gamma = 0.19626$, $t = 3.26636$, $p < 0.01$), supporting $H1$ and $H1a$. Reactive social media market orientation positively and strongly affects innovation capacity ($\gamma = 0.60888$, $t = 10.97465$, $p < 0.01$), while it seems not to directly affect big data knowledge management ($\gamma = -0.00093$, $t = -0.01916$), supporting $H2a$ and rejecting $H2$. Social media customer collaboration exhibits a positive and significant influence on big data knowledge management ($\gamma = 0.25015$, $t = 4.39176$, $p < 0.01$) but not on innovation capacity ($\gamma = 0.08382$, $t = 1.59638$), supporting $H3$ and rejecting $H3a$. Furthermore, the relationship between big data knowledge management and innovation capacity is positive and significant ($\beta = 0.16058$, $t = 2.89399$, $p < 0.01$), supporting $H4$. With regard to business performance, innovation capacity positively and significantly affects customer relationship performance ($\beta = 0.61503$, $t = 10.78117$, $p < 0.01$), while it seems not to directly affect financial performance ($\beta = -0.06812$, $t = -1.46505$), supporting $H5$ and rejecting $H6$. Finally, customer relationship performance positively and strongly affects financial performance ($\beta = 0.90387$, $t = 17.05798$, $p < 0.01$), supporting $H7$.

The structural model explains 43.08 percent of the variance in big data knowledge management ($R^2 = 0.43084$), 69.08 percent of that in innovation capacity ($R^2 = 0.69082$), 37.83 percent of that in customer relationship performance ($R^2 = 0.37827$) and 74.59 percent of that in financial performance ($R^2 = 0.74588$).

The causal relationships among constructs and the hypotheses test are synthesized in Table III.

### 4.3 Mediation effects of big data knowledge management

Table IV presents the total, direct and indirect effects from the mediation analyses and indicates the mediation types.
Proactive social media market orientation has a positive total effect on innovation capacity ($\gamma = 0.26324$, $t = 4.99144$, $p < 0.01$). It has a positive direct effect ($\gamma = 0.19626$, $t = 3.26636$, $p < 0.01$), as well as an indirect effect through big data knowledge management ($\gamma = 0.07428$, $t = 2.74181$, $p < 0.01$). This indicates partial mediation, supporting the hypothesis that big data knowledge management mediates the positive relationship between proactive social media market orientation and innovation capacity. Reactive social media market orientation has a positive total effect on innovation capacity ($\gamma = 0.48973$, $t = 10.87559$, $p < 0.01$). It has only a positive direct effect ($\gamma = 0.60888$, $t = 10.97465$, $p < 0.01$), while the indirect effect through big data knowledge management is not significant ($\gamma = -0.00012$, $t = -0.01915$; ns). This indicates that big data knowledge management does not mediate the positive relationship between reactive social media market orientation and innovation capacity. Finally, social media customer collaboration has a positive total effect on innovation capacity ($\gamma = 0.11839$, $t = 2.39724$, $p < 0.05$). The direct effect is not significant ($\gamma = 0.08382$, $t = 1.59638$). However, the indirect effect through big data knowledge management is significant ($\gamma = 0.03835$, $t = 2.41712$, $p < 0.05$). This indicates full mediation, supporting the hypothesis that big data knowledge management mediates the positive relationship between social media customer collaboration and innovation capacity.

5. Discussion
Despite the current hype surrounding big data and social media has attracted the interest of both academics and practitioners, the impact of big data acquired from social media on knowledge management practices, innovation processes and business performance remains largely unexplored. Transforming customer-related data into meaningful information through the development of knowledge management capabilities and practices has become a critical asset for firms to boost their innovation capacity and to achieve greater economic and customer value. In this context, the conceptual model proposed and tested in this study investigates the causal relationships among social media market orientation, in terms of both proactive and reactive orientation, social media customer collaboration, big data knowledge management, innovation capacity and firms’ performance outcomes, in terms of both customer relationship performance and financial one. Moreover, the study tests the mediating role of big data knowledge management to reveal if and how transforming social media customer-related information into knowledge leads firms to improve their ability to design and implement innovative products/services that address existing and latent customers’ needs.
This study offers several theoretical contributions to the extant literature on knowledge management, social media and big data management.

The first set of findings concerns the direct effects of proactive and reactive social media market orientation and social media customer collaboration, as different ways of data acquisition from social media, on big data knowledge management and innovation capacity. Findings reveal that the different ways through which firms gather social media information have different impacts on big data knowledge management and innovation capacity, displaying some interesting results. In line with previous research (Hurley and Hult, 1998; Lado and Maydeu-Olivares, 2001), market orientation, in terms of both proactive and reactive orientation, positively and significantly affects firms’ ability to implement innovative products/services in order to better satisfy customers’ needs, suggesting that these two types of market orientation are crucial in enhancing firms’ innovation capacity.

However, even if some researchers have suggested that firms that generate and use market and customer intelligence, and integrate knowledge through customer collaboration are able to improve their knowledge management practices (Nguyen et al., 2015), this study provides mixed results. In fact, the study suggests that only proactive social media market orientation and social media customer collaboration positively influence big data knowledge management, while reactive social media market orientation seems to not represent a key capability in managing and transforming customer-related information into knowledge. This result could be due to the fact that through responsive orientation firms collect information about customers’ expressed needs (Slater and Narver, 1998) that do not need to be treated and managed with ad hoc knowledge management practices to become useful for the firms’ innovative processes.

Second, the study extends the current understanding of knowledge management practices by providing empirical support for the mediating role of big data knowledge management as a critical firms’ resource in the relationships between proactive social media market orientation and innovation capacity and between social media customer collaboration and innovation capacity. These interesting results indicate that, in contrast to reactive market orientation that affects innovation capacity only directly, these other two ways of acquiring knowledge enhance innovation capacity indirectly through the mediating role of big data knowledge management. Since innovation is a high-risk and resource consuming activity (Nguyen et al., 2015), the study reveals that big data knowledge management supports firms in transforming data into meaningful information, in improving their ability to exploit customer-related knowledge arising from social media, and in strengthening innovation capacity. In line with previous studies (Fidel et al., 2016; Lusch et al., 2007) these results are important because they highlight that, thanks to big data knowledge management practices, firms can effectively bridge the gap between discovering and understanding customers’ latent needs and developing innovation. In this perspective, this study extends the mediation variables used to explain the relationship between knowledge acquisition and innovation, pointing out that big data knowledge management represents a crucial requirement for innovation and value creation. Moreover, the study also suggests to treat proactive and reactive social media market orientation as coexisting but separate constructs (Narver et al., 2004; Ordanini and Maglio, 2009) because they have different direct and indirect effects on firms’ innovation capacity, with big data knowledge management as mediator. Finally, another contribution of the study can be found in the mixed results emerged from the analysis of the direct relationship between innovation capacity and business performances. In particular, the study reveals that firms’ capacity to implement innovation and shape organizations to successfully face the dynamic competitive environment allows firms to directly achieve greater relational outcomes and, only through these relational results, improve their financial performance. In line with previous research works (Calantone et al., 2002; Taghizadeh et al., 2018; Zeng et al., 2015), this study highlights
that innovation capacity gives firms the ability to utilize their resources to realize new products and services (or new processes or marketing activities) and to better satisfy customers’ wants and needs, enhancing firms’ customer relationship performance. However, although previous studies have pointed out also the positive and direct effect of innovation on financial performance (Bigliardi, 2013), this study reveals that innovation capacity affects financial outcomes only through customer relationship performance. This result suggests that firms’ innovation capacity alone does not directly influence financial results but if the innovation capacity is used to gain customers’ satisfaction, loyalty and retention can effectively facilitate firms’ in achieving higher financial performance. In fact, innovation capacity is a key strategic resource of firms’ overall competitiveness because it supports long-term customer relationship management, enabling firms to enhance their performance and, consequently, remain competitive (Fidel et al., 2015; Singh and Gaur, 2013).

6. Conclusions, implications and future perspectives

The objective of this study is to examine the direct and indirect effect of social media market orientation, proactive and reactive, social media customer collaboration on innovation capacity and firm performances, as well as the mediating effect of big data knowledge management. The results of the study clearly show that firms need to search for and manage customer knowledge in order to innovate and, consequently, to perform better both in term of customer relationship performance and financial performance. This study’s evidences provide guidance to practitioners who are daily engaged in managing social media and exploiting big data and customer-related information.

First, findings reveal that, even if it is crucial to use social media to collect information about customers’ needs, not all data have the same explicit informative value. Consequently, information need to be treated differently in order to better understand customers’ needs and develop innovation. This result has a significant implication for practitioners because it suggests that managers have to evaluate accurately which of the data gathered from social media have to be processed through knowledge management practices to obtain knowledge useful to improve firms’ innovation capacity.

Second, the mediating role of big data knowledge management highlights how firms’ market orientation and customer knowledge can be leveraged as a source of innovation and competitive advantage. A knowledge-management oriented firm generates and disseminates customer knowledge within the whole organization in order to innovate and better target customers’ needs. From a managerial perspective, this result emphasizes the importance of developing big data knowledge management as a unique resource that can contribute to sharpen firms’ innovation capacity and, consequently, increase their competitiveness.

Finally, analyzing the link between innovation capacity and firm performance, it has emerged that firm ability to innovate contributes to enhance customer relationship performance that, in turn, increase financial performance. In this perspective, managers should be aware that firms’ innovation capacity can significantly contribute to firms’ performance, but it not always affects financial results directly. Using innovation capacity to develop products and services that meet customers’ needs and expectations, firms have the opportunity to improve customer relationships and, by satisfying customers and securing their loyalty, firms can also achieve better financial performance.

This study has also some limitations that suggest avenues for future research. First, it investigates firms operating in Italy who invest in marketing, advertising and communication. Future research should try to include in the sample also firms that operate in other countries, especially to find out differences and commonalities with foreign markets.

Second, the study adopts subjective measures to evaluate firms’ performance and, more in detail, it relies on managers’ perceptions about firms’ financial and customer relationship performance. In order to better understand the causal relationships among
the constructs investigated in the present study, future research should develop more objective measures of these variables. Finally, the study does not consider that customer information acquired through social media can be very different because each tool has its own interaction protocol and engaging instruments. Future studies should investigate if and how big data retrieved from several social media provide different insights about customers.

References


Appendix

(1) Proactive social media market orientation (adapted from Nguyen et al., 2015):
   - Our firm helps customers to anticipate developments in the markets using social media.
   - Our firm continuously tries to discover additional needs of our customers of which they are unaware using social media.
   - Our firm innovates using social media even at the risk of accelerating our products obsolescence.
   - Our firm searches for opportunities using social media in areas where customers have difficulty in expressing their needs.

(2) Reactive social media market orientation (adapted from Nguyen et al., 2015):
   - Our firm constantly monitors our level of commitment and orientation to serving customer needs using social media.
   - Our strategy for competitive advantage is based on our understanding of customer needs using social media.
   - Our firm measures customer satisfaction systematically and frequently using social media.

(3) Social media customer collaboration (adapted from Fidel et al., 2016):
   - Our firm interacts with customers to obtain useful information for innovation using social media.
   - The intensity with which our firm interacts with customers using social media is high.
   - Our firm frequently uses social media to organize meetings with customers.
   - The number of customers with whom our firm interacts using social media is high.

(4) Big data knowledge management (adapted from Alegre et al., 2011; Fidel et al., 2016):
   - Our firm uses coding systems of big data that we have collected about our customers using social media.
   - Our firm uses internal mechanisms to promote exchange of big data/information on customers.
   - Our firm uses participatory techniques among our employees and customers (such as client meetings, client interviews for improvements etc.).
   - Our firm uses tools to ensure big data about customers reach everyone in the firm.
- Our firm has information processing systems to process big data about customers.
- Our firm uses control systems and reviews the firm’s existing information on customers.
- Our firm uses systems that allow the big data that were used in previous innovation tasks to be used in new innovation tasks.

(5) Innovation capacity (adapted from Fidel et al., 2016):
- Our firm has introduced innovative products and/or services in the last three years.
- Our firm has innovated in production processes (adoption of new technologies, improved processes) in the last three years.
- Our firm has innovated in management processes (administrative area, human resources, new departments, project management) in the last three years.
- Our firm has innovated in marketing aspects (commercialization, penetrate in new markets and/or segments, new distribution channels, new forms of communication with customers and/or suppliers, new methods or pricing strategies) in last three years.

(6) Customer relationship performance (adapted from Trainor et al., 2014; Rapp et al., 2010):
- Compared to competitors:
  – Our customers work with our firm for a long time.
  – Once we get new customers, they tend to stay with our firm.
  – Our customers are very loyal to our firm.
  – Our customers are satisfied with our firm.
  – Customer retention is very important to our firm.

(7) Financial performance (adapted from Ozkaya et al., 2015; Grissemann et al., 2013):
- Compared to competitors:
  – Our sales have grown in the past two years.
  – Our market share has grown.
  – Our profitability has increased.

Note: Respondents evaluated all the measurement items on seven-point scales ranging from 1 “Strongly disagree” to 7 “Strongly agree.”

About the authors
Roberto Chierici is Research Fellow in Marketing and Management and Temporary Lecturer in Sales and Marketing at the University of Milan-Bicocca. He obtained his PhD in Business Administration, Management and Territorial Economics from the University of Milan-Bicocca in 2016. His research interests focus on different topics, such as social media and customer relationship management, social commerce, multichannel customer management, social media, information and knowledge management, and new forms of entrepreneurship. Roberto Chierici is the corresponding author and can be contacted at: roberto.chierici@unimib.it

Alice Mazzucchelli is Research Fellow in Marketing and Management at the University of Milan-Bicocca. She was awarded her PhD in Business Administration, Management and Territorial Economics at the University of Milan-Bicocca in 2016. Actually, she is Temporary Lecturer in Marketing at the University of Milan-Bicocca. She has a wide range of research interests, which tend to be multi-disciplinary in approach such as knowledge management, social commerce, social media, multichannel customer management, entrepreneurship, internationalization and enterprises bankruptcy.

Alexeis Garcia-Perez is Senior Fellow of the UK Higher Education Academy. His original background in computer science was complemented by PhD in knowledge management from Cranfield University, UK. A socio-technical understanding of information systems has enabled Alexeis to focus on the wider challenges of data, information and knowledge management in organizations and society. He has collaborated extensively with key industry players including Siemens, General Electric and the British Railway Industry. He is Visiting Research Scholar at the Georgetown University, USA.
In addition to leading research on data in organizations and society at the Research Centre for Business in Society at Coventry University, Alexeis has been the Course Director for programmes including the MBA Cyber Security.

Professor Demetris Vrontis studied in the UK and obtained a BSc (Hons) in Business from Manchester Metropolitan University Business School, a PGCE (HE) from Manchester Metropolitan University, an MBA (with Distinction) from the University of Hull and PhD in International Marketing from Manchester Metropolitan University Business School. Professor Vrontis is Fellow Member and certified Chartered Marketer of the Chartered Institute of Marketing (UK) and a Chartered Business and Chartered Marketing Consultant certified by the Chartered Association of Business Administrators (Canada), serving as a consultant and member of Board of Directors to a number of international companies. Professor Demetris Vrontis is Professor in Marketing and Dean of the School of Business at the University of Nicosia in Cyprus. His prime research interests are on strategic marketing planning, branding, international marketing and marketing communications, areas that he has widely published in over 65 refereed journal articles, contributed chapters and cases in books/edited books and presented papers to conferences on a global basis. Professor Vrontis is also the author of ten books mainly in the areas of international marketing and marketing planning.

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Purpose – Big data analytics (BDA) guarantees that data may be analysed and categorised into useful information for businesses and transformed into big data related-knowledge and efficient decision-making processes, thereby improving performance. However, the management of the knowledge generated from the BDA as well as its integration and combination with firm knowledge have scarcely been investigated, despite an emergent need of a structured and integrated approach. The paper aims to discuss these issues.

Design/methodology/approach – Through an empirical analysis based on structural equation modelling with data collected from 88 Italian SMEs, the authors tested if BDA capabilities have a positive impact on firm performances, as well as the mediator effect of knowledge management (KM) on this relationship.

Findings – The findings of this paper show that firms that developed more BDA capabilities than others, both technological and managerial, increased their performances and that KM orientation plays a significant role in amplifying the effect of BDA capabilities.

Originality/value – BDA has the potential to change the way firms compete through better understanding, processing, and exploiting of huge amounts of data coming from different internal and external sources and processes. Some managerial and theoretical implications are proposed and discussed in light of the emergence of this new phenomenon.

Keywords Performance, Big data, SMEs, Big data analytics, Knowledge management

Introduction

Nowadays, the advantages gained by managing big data can be applied by technological, digital, or online businesses. Companies that were born digital, such as Google and Amazon, are already using big data, because in the innovation process they found fewer difficulties mastering big data. But the potential to gain competitive advantage may be even greater for other companies. In fact, big data enables managers to make decisions based on evidence rather than intuition (McAfee and Brynjolfsson, 2012). Especially thanks to the drop in costs regarding all the elements of storage, memory, processing, band width, and so on, this also means that companies not born digital are able to approach big data analysis in a more economical way. So, big data has the potential to transform traditional businesses as well, as long as the technology needed to collect the big amount of data is available and cheaper than before. In this context, there has been growing use of, and reliance on, big data in a variety of
different industries and commercial contexts, from finance and health care to supply chain domains (Chae, 2015; George et al., 2014).

McAfee and Brynjolfsson (2012) found that companies that are more data-driven characterised themselves with better performance on objective measures of financial and operational results. In particular, companies in the top third of their industry in the use of data-driven decision-making were, on average, 5 per cent more productive and 6 per cent more profitable than their competitors.

Big data analytics (BDA) is the process needed to comprehend the conglomerate of data in order to extract and generate useful information and knowledge (Chen et al., 2012), which, through interpretation and categorisation, lead to more effective management (Chen et al., 2012; Davenport, 2013). The main objective of the process of collecting and analysing big data is to develop actionable insights and new knowledge to establish competitive advantages. Thus, BDA is becoming a major differentiator between high performance and low performance, as it allows firms to have a long-term vision, decrease customer acquisition costs by 47 per cent, and raise firm revenue by about 8 per cent (Liu, 2014). Thus, managers can use big data to know more about their businesses and transform the knowledge generated into efficient decisions, improving performance, and the entire decision-making process (Gupta and George, 2016). However, the management of the knowledge generated from the BDA as well as its integration and combination with firm knowledge need a structured and integrated approach.

In all the innovation activities carried out by firms, the management of relevant knowledge that lies outside or inside traditional firms’ boundaries has been addressed as one of the key managerial challenges that managers need to face, based on prior personal background and skills (Nuruzzaman, Gaur and Sambharya, 2018). Therefore, firms that are more experienced in managing different kinds of knowledge tend to be more innovative (Andersson et al., 2015; Nuruzzaman, Singh and Pattnaik, 2018) and better exploit and leverage other internal capabilities, such as BDA. In fact, the propensity and the orientation of the firm to manage and integrate new and old knowledge can improve the positive effect of BDA on firm performance (Khan and Vorley, 2017; Tian, 2017).

Moreover, the sometimes counterintuitive results of studies on the link between information technology investment and firm performance may be related to different factors about data and knowledge management (KM), such as the unavailability of appropriate data, scarce reliability of data, heterogeneity of data sources, absence of an assessment of the indirect benefits of IT, and lack of an effective KM orientation of the organisation (Brynjolfsson and Hitt, 2000; Irani, 2002; Anand et al., 2013). In this specific context of analysis, the challenges in the management of relevant knowledge coming from BDA are significant, such as data integration complexities, lack of skilled personnel, data security and privacy issues, and inadequate IT infrastructure (Gandomi and Haider, 2015).

The objective of this work is twofold. First, it empirically tests if BDA capabilities (i.e. the big data analytical capabilities from both a technological and managerial perspective) have a positive impact on firm performance; second, it aims at understating if KM plays an important role in determining the outcomes of BDA capabilities (i.e. it has a mediator effect). An empirical analysis was developed using structural equation modelling with data collected from 88 Italian SMEs. Findings show that firms that developed more BDA capabilities than others increased their performances and that KM orientation plays a significant role in mediating the effects of BDA capabilities on firm financial performances of the Italian SMEs in our sample. The remainder of the paper is organised as follows: first, relevant theories are presented along with the main hypothesis, and then the methodology of the paper is explained. After this, findings and conclusions are presented, highlighting the main contributions, limitations, and future lines of research.
Theoretical background

Big data

The concept of big data is defined by the literature as huge amounts of structured and unstructured data, accessible in real time (Einav and Levin, 2013; O’Leary, 2013). These kinds of data are available in nature, as data are everywhere, but because of their complexity, they cannot be processed using traditional methods (Mackenzie, 2006). In recent years, due to these advantages, the phenomenon of big data has gained interest in academia and business, as both acknowledge its high operational and strategic potential in generating business value.

Big data can be defined as the approach to manage, process, and analyse five dimensions of data. In the literature, they are known as the “5 Vs”:

1. **Volume.** With the never-ending technological innovation, the quantity of data created every day grows exponentially. Every second on the internet the amount of data generated is more than the storage capability of the entire internet of 20 years ago.

2. **Variety.** The sources of big data are many and relatively new. In fact, data are generated from different digital platforms. Through their inputs on the numerous digital devices, consumers provide information about their habits, needs, and desires. For example, big data can take the form of messages, updates, images posted on social networks, readings from sensors, GPS signals from cell phones, and more.

3. **Velocity.** The speed of data creation is even more important than the volume, because the economic world is becoming more and more competitive, and one of the key factors for success is the ability to make decisions faster. Today, data are obtainable in real time or nearly real time, and this makes it possible for a company to be much more agile and faster in the decision-making process.

4. **Veracity.** The data collected must have quality, and the original source must have a certain level of trust. Veracity refers to the fact that data may contain noise or be incomplete and out-of-date.

5. **Value.** Extracting economic benefits from the available big data has enormous importance. This value is often linked to the ability of the organisation to make better decisions.

In terms of technology, big data still has a long way to go before it lives up to the claims currently being made for it. Currently, three technologies are central to the future of big data (Janssen et al., 2017; White, 2012). The Hadoop open source framework: processing, storing, and analysing massive amounts of distributed, unstructured data; NoSQL databases: serving up discrete data stored among large volumes of multi-structured data to end-user and automated big data applications; and massively parallel analytic databases: using massively parallel processing that allows for the ingesting, processing, and querying of data on multiple machines simultaneously.

According to Kauffman et al. (2012, p. 85), the concept of big data is skyrocketing in particular “due to social networking, the internet, mobile telephony and all kinds of new technologies that create and capture data”. Thus, the relevance of BDA is growing and will continue to grow in the next decade. The next challenge will be how to make effective use of all these data and how to create value from them, particularly in emerging contexts such as smart cities (e.g. Scuotto et al., 2016). Big data can enhance a firm’s financial performances through a better destination of resources (Bresciani et al., 2017). However, big data is sometimes mistaken for analytics, but the big data movement has the objective of draining intelligence from data and translating it into a business advantage.
Big data analytics

BDA is emerging as a hot topic among scholars and managers, and it has been referred to as the capacity of firms to manage, process, and analyse big data (Wamba et al., 2017). An emerging stream of literature is developing quickly, highlighting the positive effects of BDA within organisations. For example, Hagel (2015) showed how BDA is increasingly becoming a key component of decision-making processes in different kinds of businesses due to a new proactive and forward-looking approach. Wills (2014) pointed out that almost 35 per cent of purchases made on Amazon.com are generated from personalised purchase recommendations to customers based on BDA.

However, the value extracted from data does not depend only on the quality of the data themselves but also on the quality of the different processes in which data are collected and analysed. This often requires multiple actors from different disciplines and diverse processes and practices (Ferraris et al., 2016; Janssen et al., 2017). In order to obtain the full benefits from big data, managers need to align existing organisational culture and capabilities across the whole organisation. Barton and Court (2012) highlighted that the key challenge for using big data is to make big data trustworthy and understandable to all employees. They exemplified that frontline employees in a retail industry were reluctant to use big data since they either did not rely on a big data-based model or were not capable of understanding how it works. Shah et al. (2012), for example, opined that business analytics skills are still confined to the “expert” level and are not yet disseminated to all in an organisation; however, in order to add value from using big data, it is essential that all levels of employees are well equipped about big data, which can be achieved through training. Indeed, return on investment in big data would not materialise unless employees at all levels are able to understand and include data in their decision-making (Shah et al., 2012).

Therefore, the new decision-making process brought about by analysing big data leads to a change in how decisions are made and who makes them (McAfee and Brynjolfsson, 2012). Until now, without taking into consideration the information provided by big data, it made sense to give the responsibility to well-placed people, who usually make decisions based on experience acquired through their career and patterns and relationships they have observed.

Facing the change of big data, the person in charge of the decisions and the so-called highest paid person opinions will decrease in their value because of the analytic information of big data. In short, managers could target more effective interventions in areas that so far have been dominated by gut and intuition rather than by data and rigour. Executive managers can make decisions based on big data by getting into the habit of asking, “What do the data say?” Even knowing this, some analysts estimate that only two in three business leaders trust the information they use to make decisions (McAfee and Brynjolfsson, 2012). Therefore, “if data is not of sufficient quality by the time it has been integrated with other data and information, a false correlation could result in the organization making an incorrect analysis of a business opportunity” (White, 2012, p. 211).

In line with Wamba et al. (2017) and with the IT capabilities literature (e.g. Gupta and George, 2016), we saw BDA capability as an important organisational capability leading to sustainable competitive advantage in the big data environment.

Hypothesis development

Yiu (2012) highlights that BDA allows for improved data-driven decision-making and innovative ways to organise, learn, and innovate, resulting in the better management of different firm processes (e.g. CRM, operational risks, and production efficiency) that lead to better performances.

Past research on information system (IS) and resource-based view (RBV) showed that the organisational capability to process information have a positive effect on firm
performance, such as return on investment or higher profitability (e.g. Wade and Hulland, 2004; Contractor et al., 2016). Organisational capabilities may include skills and routines in transforming knowledge inputs into value outputs of greater value (Ferraris et al., 2017).

So, the ability of an organisation in collecting, preparing, and analysing big data may make the difference, in particular if the organisation make these processes difficult to imitate, thanks to idiosyncrasies or to path dependency effects (e.g. Customised infrastructure) (Dierickx and Cool, 1989; Janssen et al., 2017).

In order to achieve a competitive advantage, firms need to combine and deploy several firm-level resources and capabilities. In this case, big data alone is not sufficient to create BDA capabilities (Gupta and George, 2016; Wamba et al., 2017). A complex mix and interrelation among different financial, human, physical and organisational resources are thus essential to create these superior, rare, and difficult-to-imitate capabilities, which, in turn, lead to better performances, such as better customer retention or return on investment and increase in sales growth and profitability (Giacosa et al., 2018). Thus, we propose that:

**H1.** The greater the firm’s BDA capabilities, the higher the firm performances are.

Knowledge has been defined as a set of justified beliefs that can be arranged and managed to enhance the organisation’s performance through effective action (Nonaka, 1994; Alavi and Leidner, 2001). There are three acknowledged major KM processes: the acquisition, conversion, and application of knowledge (Alavi et al., 2006; Gasik, 2011; Gold et al., 2001). Knowledge acquisition is the process used to develop new knowledge from data and information, while knowledge conversion refers to making the acquired knowledge useful for the organisation (Gold et al., 2001) by structuring it or transforming tacit knowledge into explicit knowledge. Knowledge application refers to the use of knowledge to perform tasks (Sabherwal and Sabherwal, 2005).

Thus, KM includes the firm’s processes of acquiring new knowledge, converting knowledge into a form that is usable and easily accessed, and applying this knowledge in the organisation (Gasik, 2011), affecting firm performances. KM processes allow companies to capture, store, and transfer knowledge efficiently (Magnier-Watanabe and Senoo, 2010) in order to improve customer satisfaction, market share, and financial results. Therefore, we propose the following:

**H2.** The greater the KM orientation, the higher the firm performances are.

Prior studies found out that part of the business value derived from IS investments have reported mixed results, resulting in the so-called “IT productive paradox” (Gupta and George, 2016). Indeed, some scholars have argued that IS and KM investments do not necessarily lead to improved operational efficiency and effectiveness (e.g. Irani, 2010) while others identified a positive association between these investments and firm performance (Ferraris et al, 2018). Their findings suggest that the absence of a positive link between IS and KM investment and firm performance found by prior studies may be explained by several factors including the unavailability of appropriate data, the existence of time lags between these investments and the business value generated from these investments, the absence of an assessment of the indirect benefits of IT, and a lack of an effective KM orientation of the organisation (Anand et al., 2013).

Boyd and Crawford (2012, p. 6) underlined that BDA “reframes key questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and the categorisation of reality”. Firms usually face challenges in the management of relevant knowledge coming from BDA (e.g. data integration complexities, lack of skilled personal, data security and privacy issues, and inadequate IT infrastructure; Gandomi and Haider, 2015), and a structured approach to KM is even more important in this context.
In fact, BDA could bring changes in personal KM, increasing and widening the role of individual knowledge and changing the role of workers (Pauleen, 2009). Recently, Khan and Vorley (2017) highlighted the significant interdependence between BDA and KM in order to help not only in the sharing of common knowledge of business intelligence but also in extending human knowledge, resulting in different kinds of firm performance improvements. Thus, we propose the following:

**H3. KM acts as a mediating variable between a firm’s BDA capabilities and firm performances.**

**Research methods**
Based on the RBV theory, we conceptualised the research model, developed the survey, and validated the hypothesised relationships using the elliptically reweighted least square method as an estimation procedure (Tippins and Sohi, 2003; Alegre et al., 2013).

The administration of the survey took place in three stages (Darroch, 2005). First, we identified a total of 159 Italian firms that are members of a national association for BDA. Second, a pre-notification letter was sent to all firms’ CEOs explaining the objective of the survey, announcing the arrival of the questionnaire, and asking them to direct the questionnaire to the person who should be more confident to answer on this topic within the organisation. Third, three weeks later, a copy of the survey was mailed with a cover letter. The usable sample size was 88, and the effective response rate was 55.3 per cent.

All the firms in our sample are SMEs according to the definition of the Organisation for Economic Co-operation and Development (OECD, 2005), characterised by a number of employees up to 249 and a turnover of up to 50m euro. More specifically, SMEs participating in our analysis have employees ranging from about 80 to 230, and revenues ranging from 28 to 45m euro.

The questionnaire was built according to the mainstream literature, highlighted in the section below. We also limited the possibility of common method variance by separating the questions and items in the questionnaire, limiting the risk of respondents’ rationalising their answers and establishing methodological separation of measurement (Podsakoff et al., 2003).

We addressed the psychometric properties of the measurement scales following prior studies and accepted practices (Anderson and Gerbing, 1988; Tippins and Sohi, 2003; Alegre et al., 2013). They included content validity, reliability, and discriminant and convergent validity (see Table I for means, standard deviations, factor correlations, and reliability indexes for the main constructs). To address content validity, we built the scales based on the relevant previous literature. In line with Alegre et al. (2013), we assessed reliability using Cronbach’s α coefficient reporting and composite reliabilities. Reliabilities for the dimensions of our latent constructs appeared to be adequate (see Table I). Confirmatory factor analysis (CFA) was used to address discriminant and convergent validity by comparing the $\chi^2$ differences between a constrained confirmatory factor model (where the correlation between two factors is set to 1, indicating they are the same construct) and an unconstrained model (where the correlation between two factors was free). All $\chi^2$ differences were significant, confirming the discriminant validity (Anderson and Gerbing, 1988). While constraining the confirmatory factor model to 0, the same procedure was used to obtain further support for convergent validity. Finally, we performed a CFA to confirm that all scale items loaded significantly on their hypothesised construct factors (Anderson and Gerbing, 1988), confirming convergent validity.

**Variables**
In order to build our dependent variables (firm performance), based on some previous studies (Tippins and Sohi, 2003; Wang et al., 2012), we asked our respondents the extent to
Using analytics improved _____ during the last three years relative to competitors with regard to customer retention, sales growth, profitability, and return on investment. We used a seven-point Likert-type scale ranging from 1 to 7. The average value was used when constructing the variable firm performance (Cronbach’s α = 0.81).

Regarding BDA capabilities, we followed previous studies (e.g. Akter et al., 2016) to assess the management and technological aspects of BDA. First, regarding the BDA management capabilities, six factors were used (Kim et al., 2012): BDAM1. We continuously examine the innovative opportunities for the strategic use of BDA; BDAM2. We perform BDA planning processes in systematic and formalised ways; BDAM3. When we make BDA investment decisions, we think about and estimate the effect they will have on the productivity of the employees’ work; BDAM4. When we make BDA investment decisions, we consider and project about how much these options will help end-users make quicker decisions; BDAM5. In our organisation, information is widely shared between business analysts and line peoples, or those who make decisions or perform jobs have access to all available know-how; BDAM6. In our organisation, the responsibility for BDA development is clear. Second, six factors were used for BDA technological capabilities (Byrd, 2000): BDAT1. All remote, branch, and mobile offices are connected to the central office for analytics; BDAT2. Our organisation utilises open system network mechanisms to boost analytics connectivity; BDAT3. Software applications can be easily transported and used across multiple analytics platforms; BDAT4. Our user interfaces provide transparent access to all platforms and applications; BDAT5. End-users utilise object-oriented tools to create their own analytics applications; BDAT6. Applications can be adapted to meet a variety of needs during analytics tasks. The average value was used when constructing the variable BDA capabilities (Cronbach’s α = 0.85). Due to the fact that these firms are members of big data associations and defined themselves as pioneers in the use of BDA, we decided not to include items regarding the talent of the workforce and human capital in order to avoid bias in the answers.

To develop the variable KM orientation, we followed Darroch (2005) according to three components of KM: knowledge acquisition, knowledge dissemination, and responsiveness to knowledge. Regarding the knowledge acquisition scale, we used different items: KAF1. Organisation values employees’ attitudes and opinions; KAF2. Organisation has well-developed financial reporting systems; KAF3. Organisation is sensitive to information about changes in the marketplace; KAF4. Science and technology human

<table>
<thead>
<tr>
<th>Factors</th>
<th>Composite reliability</th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) BDAM capabilities</td>
<td>0.79</td>
<td>3.126</td>
<td>0.818</td>
<td>(0.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) BDAT capabilities</td>
<td>0.81</td>
<td>3.055</td>
<td>0.736</td>
<td>0.713**</td>
<td>(0.87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Knowledge acquisition</td>
<td>0.83</td>
<td>3.653</td>
<td>0.796</td>
<td>0.454**</td>
<td>0.406**</td>
<td>(0.85)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Knowledge dissemination</td>
<td>0.82</td>
<td>3.823</td>
<td>0.787</td>
<td>0.288**</td>
<td>0.294**</td>
<td>0.615**</td>
<td>(0.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) Responsiveness to knowledge</td>
<td>0.79</td>
<td>3.024</td>
<td>0.816</td>
<td>0.227*</td>
<td>0.274**</td>
<td>0.613**</td>
<td>0.604**</td>
<td>(0.82)</td>
<td></td>
</tr>
<tr>
<td>(6) Firm performance</td>
<td>0.83</td>
<td>4.206</td>
<td>0.799</td>
<td>0.213**</td>
<td>0.354**</td>
<td>0.352**</td>
<td>0.365**</td>
<td>0.512**</td>
<td>(0.81)</td>
</tr>
<tr>
<td>Concepts</td>
<td>Mean</td>
<td>SD</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) BDA capabilities</td>
<td>3.06</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) KM orientation</td>
<td>3.45</td>
<td>0.78</td>
<td>0.59**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Firm performance</td>
<td>4.23</td>
<td>0.81</td>
<td>0.35**</td>
<td>0.70*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** n = 88. α reliabilities are shown on the diagonal. *p < 0.05; **p < 0.01
capital profile; KAF5. Organisation works in partnership with international customers; KAF6. Organisation gets information from market surveys. Items were measured on a five-point Likert-type scale ranging from 1 (very low) to 5 (very high). Regarding the knowledge dissemination scale, we used different items: KDF1. Market information is freely disseminated; KDF2. Knowledge is disseminated on the job; KDF3. Use of specific techniques to disseminate knowledge; KDF4. Organisation uses technology to disseminate knowledge; KDF5. Organisation prefers written communication. Finally, the knowledge responsiveness scale was built using the following items: KRF1. Responds to customers; KRF2. Well-developed marketing function; KRF3. Responds to technology; KRF4. Responds to competitors; KRF5. Organisation is flexible and opportunistic. The average value was used when constructing the variable KM orientation (Cronbach's $\alpha = 0.89$).

All the items were measured using a seven-point Likert scale. The correlations, means, standard deviations, and reliability indexes for the main constructs are provided in Table I.

In line with prior studies (e.g. Alegre et al., 2013), we decided to include two variables in the overall model because they may impact firms' performance: firm size and R&D expenditure. Both may influence the effective use of big data and the related capabilities, as well as the capacity to understand knowledge because, as suggested by Cohen and Levinthal (1990), individuals learn, or absorb, knowledge by associating it with their existing knowledge base. First, we measured firm size through a logarithmic transformation of two indicators: turnover and the total number of employees. Second, we measured R&D through a logarithmic transformation of two indicators: the R&D budget and the number of R&D employees (Alegre et al., 2013).

**Findings**

We tested for the presence of a mediating effect by performing a competing model analysis (i.e. two substantive models were estimated and evaluated for significant differences). The first model (direct effect) examined the direct relationship between BDA capabilities (BDAC) and firm performances. This model was used to test $H1$. A second model (partial mediation) examined the same relationship with firms' KM orientation, acting as a mediator. Following Tippins and Sohi (2003), the mediating effect of KM is supported when the variance explained in firm performance by the second model is higher than in the direct model; the relationship between BDAC and KM orientation is significant; the significant relationship observed in the direct model between BDAC and firm performance is greatly diminished or eliminated in the second model; and the relationship between KM orientation and firm performance is significant.

The findings of this research are significantly related to the interrelations of BDA and KM capabilities that showed how they lead to greater performances. Overall, relevant fit indexes showed good fit (Tippins and Sohi, 2003; Alegre et al., 2013). In Figure 1, all the hypothesis are confirmed, highlighting a positive association between BDAC and firm performance ($\beta = 0.59, \ t = 3.84, \ p < 0.01$) and a mediation effect of KM orientation on this relationship.

First, the variance explained in firm performance by the second model is higher than the first one (0.57 vs 0.34). Second, the relationship between BDAC and KM orientation is significant ($\beta = 0.73, \ t = 4.54, \ p < 0.01$), as well as the one between KM orientation and firm performance ($\beta = 0.78, \ t = 4.29, \ p < 0.01$). Third, the significant relationship observed in the direct model between BDAC and firm performance becomes non-significant in the second model ($\beta = 0.11, \ t = 0.159$), and a significant relationship between KM orientation and firm performance is confirmed. Together, these points provide evidence that there exists a discernible mediating effect of KM on the relationship between BDAC and firm performance, thus confirming the relevance of BDA combined with proper management of relevant knowledge.
Conclusion

Discussion and contributions

One of the most powerful aspects of the big data revolution is the unification of large data sets with advanced analytics for problem solving. This ability to spot patterns and solve problems beyond human mental capabilities has led to two main sources of insight derived from big data. First, very large and multidimensional data sets can be examined to look for previously hidden patterns and correlations. Sometimes this can validate positions that were previously supported by common sense, practical experience, or received wisdom. On other occasions, this sort of analysis can deliver entirely new insights into the underlying dynamics of a population, market, or business. Second, big data opens up the realm of reliable predictive analytics. By examining the relationships embedded in large data sets, it is possible to build a new generation of models describing how things are likely to evolve in the future.

Organisations are increasingly looking to big data in order to improve different kinds of performances (Akter et al., 2016). In this paper, we empirically found that to fully gain the benefits of big data, firms must possess BDA capabilities and a certain level of KM orientation. This may lead to a better decision-making process locating information and the relevant decision levels in the same place (Shah et al., 2012). Therefore, in the big data era, information is created and transferred, and expertise is often not where it used to be. The skilled leader may thus create an organisation flexible enough to minimise the “not invented here” syndrome and maximise cross-functional cooperation (Shams et al., 2018). People who understand the problems need to be brought together with the right data, but also with the people who have problem-solving techniques that can effectively exploit them.

The decisions made while taking into consideration big data tend to be better decisions and lead to better performance. Leaders should understand this and build BDA-specific capabilities in their workforces or be replaced by others who are able to do so. In sector after sector, companies that figure out how to combine domain expertise with data science will
have greater competitive power than their rivals. However, from this research, it has been underlined how the development of KM capabilities may enhance these positive effects, leading to greater performance.

Practically speaking, this approach can be combined with scenario planning to develop a series of predictions for how a system will respond to different policy choices. The state of the art in predictive analytics can deliver forecasts for some domains with a very high degree of precision, providing an auditable, scientific basis for making decisions in complex systems. Using BDA for problem solving has other advantages beyond seeing deep and far. Increased use of computational techniques can free up an organisation’s staff to focus on tasks where human beings continue to outperform computers, increasing overall productivity. Besides other challenges related more to IT-related aspects, the main managerial challenge is to redefine the understanding of “judgement of information”, facilitating a shift in the organisation’s decision-making approach and in organisational culture. Thus, companies must not only hire scientists who can find patterns in data and translate big data into useful business information in order to have success but also change the managerial mindset, re-orienting it to a more digital- and data-driven culture.

This paper has at least three theoretical contributions. First, it empirically validates the impact of BDA on firm performances, confirming the emergent relevance of these IT-related capabilities to the overall competitiveness of modern and more “digital” firms (Janssen et al., 2017; Wamba et al., 2017). Second, it contributes to the stream of literature on KM, highlighting another positive effect of developing KM capabilities and orientation. These capabilities have the potential to amplify the positive effects of BDA on firm performances, reducing the risks associated with big data management through systematic selection and dynamic management of relevant knowledge that is both internal and external to the firm (George et al., 2014; Janssen et al., 2017). Third, this study may contribute to the dynamic capability stream of literature by addressing the role of two important capabilities and their positive interconnections. This may be complemented with prior studies that argued about the positive link between that IT capability and firm outcomes with dynamic capabilities of the business process that mediates the relationship (Chen et al., 2012) in order to better understand the sources of IT-based competitive advantage (Wamba et al., 2017).

Concluding discussion and future line of research
While most of the articles on big data have been primarily written by technology consultants (or vendors), the focus has been still remained on the understanding of the technology along with the mainly definition and characteristics of big data. So, little knowledge about how organisations build BDA capabilities has been developed since the emergence of this phenomenon. The main aim of this paper is to shed lights on the relevance of the underlining capabilities useful to fully take advantage from big data. Considering bid data as a resource (Gupta and George, 2016), we can argue that competitive advantage may be created and sustained through the building and combining of new capabilities related to the use of big data. In fact, investments in big data technologies are not enough to create value and to build a new source of competitive advantage. A firm thus needs a unique bundle of resources in order to create capabilities superior to the competitors, that are also dynamic in relation to the change of the external environment and difficult to be imitated (McAfee and Brynjolfsson, 2012). In this paper, we do not argue that the characteristics of big data are not important but we focus the attention on the creation of value exploiting big data technology, this in turn leads to a greater performance of the firms that develop particular kind of capabilities.

In particular, we pose the attention on two key capabilities related to the human side of big data: BDA and KM capabilities. Future studies may focus on other key resources useful for the exploitation of big data, such as the development of a data-driven culture that shift the attention of the managers on the way they take decisions (from intuition to data) or the
technical skills possessed by the employees in terms of education and trainings related to big data-specific competencies.

As all the studies, this research has some limitations. First, this study is limited to Italian SMEs only and future research may be based on different samples from different industries and different geographic contexts in order to derive wider and deeper implications. For now, we limited the conduct of the study within the specific domain of BDA and in one context. However, it is also true that BDA by its nature is context specific due to the variations in the analytics industry. Second, this study used cross-sectional data so it is suggested to retest the findings using panel data to investigate its stability across time and also to potentially evaluate a time-lag effect of the BDA and KM capabilities on future firm performances.

References


Further reading


Corresponding author

Alberto Ferraris can be contacted at: alberto.ferraris@unito.it
Big data visualisation, geographic information systems and decision making in healthcare management

Anitha Chinnaswamy
Faculty of Business and Law, Coventry Business School, Coventry University, Coventry, UK
Armando Papa
Department of Computer Science, University of Turin, Turin, Italy
Luca Dezi
Department of Economics and Law, Parthenope University of Naples, Napolì, Italy, and
Alberto Mattiacci
Department of Communication and Social Research, University of Rome Sapienza, Rome, Italy

Abstract
Purpose – The World Health Organisation estimates that 92 per cent of the world’s population does not have access to clean air. The World Bank in 2013 estimated that only air pollution (AP) was responsible for a $225bn cost in lost productivity. The purpose of this paper is to contribute to the current scholarly debate on the value of Big Data for effective healthcare management. Its focus on cardiovascular disease (CVD) in developing countries, a major cause of disability and premature death and a subject of increasing research in recent years, makes this research particularly valuable.

Design/methodology/approach – In order to assess the effects of AP on CVD in developing countries, the city of Bangalore was selected as a case study. Bangalore is one of the fastest growing economies in India, representative of the rapidly growing cities in the developing world. Demographic, AP and CVD data sets covering more than 1m historic records were obtained from governmental organisations. The spatial analysis of such data sets allowed visualisation of the correlation between the demographics of the city, the levels of pollution and deaths caused by CVDs, thus informing decision making in several sectors and at different levels.

Findings – Although there is increasing concern in councils and other responsible governmental agencies, resources required to monitor and address the challenges of pollution are limited due to the high costs involved. This research shows that with developments in the domains of Big Data, Internet of Things and smart cities, opportunities to monitor pollution result in high volumes of data. Existing technologies for data analytics can empower decision makers and even the public with knowledge on pollution. This paper has demonstrated a methodological approach for the collection and visual representation of Big Data sets allowing for an understanding of the spread of CVDs across the city of Bangalore, enabling different stakeholders to query the data sets and reveal specific statistics of key hotspots where action is required.

Originality/value – This research has been conducted to demonstrate the value of Big Data in generating a strategic knowledge-driven decision-support system to provide focused and targeted interventions for environmental health management. This case study research is based on the use of a geographic information system for the visualisation of a Big Data set collected from Bangalore, a region in India seriously affected by pollution.

Keywords Big data, Knowledge management, Decision making, Smart city, Healthcare management

Paper type Research paper

1. Introduction
1.1 Cardiovascular diseases
Cardiovascular Diseases (CVDs) are the world’s leading cause of death. Approximately 80 per cent of all cardiovascular-related deaths, however, occur in low- and middle-income countries and at a younger age in comparison to high-income countries (Gersh et al., 2010).
Countries in Asia such as China and India, particularly, are burdened by the disease. In India, 41 per cent of urban male deaths and 37 per cent of urban female deaths were reported to be due to CVDs (Celermajer et al., 2012).

The main determinants of CVD, until recently, included lifestyle factors, diet, health history, hereditary factors, smoking and alcohol consumption. Over the last decade, studies have emerged, particularly in the developed nations that include air pollution (AP) as a determinant for CVD. AP is increasingly becoming a global concern and is believed to be amongst the leading causes of death in the world today. The World Health Organisation (WHO) estimates that 2m people die prematurely due to AP (WHO, 2011a). Developed countries have recognised AP as a major public health concern and have developed strategies to effectively tackle it and improve air quality. However, AP in developing countries and particularly in Asian cities is relatively high when compared to cities of the developed world (Schwela et al., 2006). The levels of AP in these cities regularly exceed WHO recommended guidelines with smoke and dust particles being double the world average (Schwela et al., 2006). Delhi, the capital city of India, is the highest polluted city in the world (The Indian Express, 2014). The common air pollutants such as NOx, O3 and PM are a significant challenge in Asian cities. Furthermore, it has been established that 13 of the world’s top 20 polluted cities are in India, followed by cities in Pakistan and Bangladesh (Roychoudhury, 2014).

Several epidemiological studies that include long- and short-term studies have produced a positive and statistically significant association between high levels of pollution and occurrences of CVDs, most of which emerge from developed countries. Although research already exists in developed nations, it is not viable to use the findings as inferences for developing nations due to the apparent differences in the sources of pollution, living conditions, lifestyles and quality of healthcare delivery. Hence, there is a growing need to understand the magnitude of AP and the patterns of CVDs in developing countries, especially in urban populations and there is an urgent need to fill the knowledge gap of the impact of AP and CVD in Asia. This research aims to contribute to this gap in knowledge.

1.2 Rationale
Given the relevance of spatial location in the health and well-being of a person, this research explores the opportunities provided by technologies to generate and use a Big Data set that allows for a spatiotemporal methodology to assess the effects of AP on the health of a population, particularly in its relation to CVDs. Based on a Geographical Information System (GIS) an environmental health information system (ENVHIS) is developed that will assist in using Big Data to inform the relevant stakeholders in addressing the challenges posed by AP and CVD in cities.

The rest of this paper is structured as follows: an overview of theories supporting this research is provided in Section 2, covering both CVDs and technology developments. This is followed by a description in Section 3 of the methodology used to conduct the research, which included development of geographic information systems for the use of Big Data from a case study of Bangalore, India. Section 4 outlines the findings of the research. Conclusions on the value of Big Data for decision making in pollution and healthcare management are presented in Section 5, which also includes an outline of areas for future developments in this area.

2. Literature review
2.1 Big Data and applications in healthcare
In recent years, the technology landscape has continuously evolved resulting in new business trends and reshaping of competitive strategies across a wide range of industries. Information technologies and Big Data are changing not only the communication landscape
but also resulted in the real world becoming smaller and more connected. Hence, companies are continuously seeking new opportunities to benefit from data.

The concept of Big Data has a broad category of applications based on the organisational capability of collecting and storing enormous quantities of information that has been derived from various internal and external sources or, alternatively, by purchasing it from specialised operators. Currently, the biggest emerging technology trends include different kinds of data management tools and smart devices that not only collect and transfer data, but also analyse it in real-time.

Big Data analytics, IoT solutions and smart technologies are about developing innovative ways to access, discover and store new knowledge in fields traditionally considered static, anchored to the traditional competition paradigms, by imposing to the operators to change their value proposition, innovating with new and dynamic solutions.

Big Data is growing in popularity, with organisations contributing to the phenomenal growth of data. Healthcare institutions in particular face many challenges ranging from disease outbreaks to treatments and evidence-based decisions. Big Data analytics has the capability to contribute to these challenges specifically by deriving meaningful insights to improve the care provided and targeting interventions.

Big Data technology helps to enhance operational efficiencies and add value for sustainable practices and policy making. Using a Big Data approach with GIS allows analysis and decision making from huge data sets, by using algorithms, query processing and spatiotemporal data mining. GIS tools for Big Data processing facilitate deep insights and predictive modelling for policy making in healthcare. GIS allows the integration of data sets either structured or unstructured and facilitates spatial analysis. The spatiotemporal queries of the big geospatial data enable better understanding of spatial trends and relationships (GIS Lounge, 2014).

2.2 Economic and Societal impact of CVD

Of all the non-communicable diseases (NCD), globally, CVD contribute to the largest cause of mortality. CVDs, a group of disorders of the heart and blood vessels account for 30 per cent of all global deaths, three times more deaths than are caused by infectious diseases including HIV/AIDS, tuberculosis and malaria combined (WHO, 2009). All countries experience the human, social and economic consequences of NCDs, but this has been particularly devastating for developing countries. Low- and middle-income countries contribute to about 80 per cent of CVD deaths (WHO, 2005). The Global Status Report (WHO, 2014) estimated that during 2011–2025, the cumulative economic losses due to NCDs under a “business as usual” scenario in low- and middle-income countries have been estimated at $7 trillion. NCDs have become a major public health problem in India accounting for 62 per cent of the total burden of foregone DALYs and 53 per cent of total deaths (Thakur et al., 2011).

The epidemiologic disease transition is largely due to economic development, urbanisation and changes in social organisation within countries and regions. Migration to urban areas for better opportunities and quality of life has posed challenges such as inadequate infrastructure, water and AP and a growing health crisis. Researchers have established links between urbanisation and its effects on the occurrences of CVD due to the increased exposure to risk factors driven by changes in diet, physical activity and environment (Smith et al., 2012). Most CVDs are preceded by risk factors that are classified as modifiable (can be treated or controlled) and non-modifiable (cannot be treated or controlled).

These risk factors are conditions or habits that increase the chances of CVD occurrence. The more risk factors an individual has, the greater the chance of developing CVD. Non-modifiable risk factors include age, family history, ethnicity and gender. Modifiable risks include tobacco use, diabetes, hypertension, dyslipidaemia, unhealthy diet, sedentary lifestyles and the harmful use of alcohol.
The human, social and economic consequences of CVDs are felt by the society in large. The effects of CVD are not only a health concern but also affects the social aspects of life. As the disease correlates to a decreased life expectancy, there is a huge impact on the individual as well as their family (HealthTalk, 2014). Individuals are known to be concerned about not being able to see their children or grandchildren grow up, they also go through a range of emotions that include anxiety, frustration and depression.

Healthcare in India is highly privatised, with most of the outpatient or inpatient care sourced through the private sector (NSSO, 2006). The large population lacks any form of health cover with nearly 90 per cent of the population paying for healthcare as out-of-pocket expenditures (MOHFW, 2005). These forms of expenditures have catastrophic effects on households where an individual is affected by CVD. India is witnessing more and more younger people affected by CVD in the country, with an increasing number of them being the sole bread winners for the family. This has a devastating social and economic impact on individuals and their households.

There is a dire need to address the growing problem of CVD. CVD is a largely preventable disease; evidence has indicated that with timely effective treatment and adequate intervention and prevention measures, the suffering due to CVD can be reduced and prevented. CVD policies are to be based on the best available evidence-based approaches that result from well-conducted, systematic reviews of the relevant evidence. Hence by using a GIS system to visualise the extent of the CVDs and AP hotspots, interventions maybe focused on the high-risk populations identified.

2.3 Geographical information systems
Geographical location is an important determinant for any activity. In the health data, there is a spatial component that can be tied to a place such as an address, postcode, region or coordinated reference location. To gain a better advantage of the spatial and temporal components in decision making, the appropriate tools must be employed.

GIS is an integrated computer system that stores information about spatial and related non-spatial data (ESRI, 2012). Boulos (2004) describes GIS as a potentially powerful resource due to their ability to integrate separate sets of data from disparate sources, provide analysis to produce new information, provide mapping and visualisation techniques thus empowering decision makers in knowledge management and decision making. GIS is gaining popularity in environmental and demographic research areas and is increasingly used in health interventions for areas and populations at risk (Caley, 2004). Due to the strong capabilities of GIS they are used in a broad range of applications such as urban planning, forestry, climate science, emergency management, epidemiology and public health.

2.3.1 GIS applications in environmental health
GIS has been widely used in determining environmental exposures and effects on health. Some examples are Guthe et al. (1992) who used GIS to predict populations of children at high risk of exposure to lead in New Jersey. Wartenberg et al. (1993) identified populations at risk of exposure to magnetic fields. Glass et al. (1995) used GIS to investigate environmental risk factors for Lyme disease in Maryland, USA.

A study by Wang et al. (2011) focused on Air Quality data for Ozone and asthma hospitalisation rates between the years 2000 and 2008. The project linked different existing systems, such as environmental and disease tracking system, to assist both the environmental agencies and health professionals to identify geographical and geospatial patterns of populations at risk. The data were geocoded into a database and analysis functions such as overlay of layers were performed. Queries on the data on the basis of county, health outcome and year were displayed in maps, charts and tables. This system, according to Wang et al. (2011), has facilitated public health officials and environmentalists to increasingly use GIS technology for data presentation, analysis and in the role of decision making.
Health GIS is widely receiving attention and within the domains of environmental health, disease ecology and public health, GIS are becoming an indispensable tool for processing, analysing and visualising spatial data (Kistemann et al., 2002).

One of the significant challenges in using GIS is obtaining the relevant data for the system. Data lies at the heart of GIS and is defined as an abstraction of reality and it is crucial to decide the data that is required for the purpose. The relevant data must also be in accordance with the capabilities of the GIS. The quality of the data is very important and must adhere to data quality aspects such as accuracy, precision, time, currency and completeness. Aronoff (1989) stated that there is an inverse relationship between data quality and data cost. Although the benefits of GIS have been highlighted and its applications have been demonstrated in developed countries, the use of GIS is limited in developing countries due to diverse challenges (Ramasubramanian, 1999). These may be due to lack of infrastructure, lack of skilled human resources, lack of knowledge and awareness and the relatively large costs associated with its setup.

3. Methodology
3.1 Case study research
In order to assess the effects of AP on CVD in developing countries, a case study area is sought that is a good representative of the rapidly growing cities in the developing world. It is essential that the selected case reflects the characteristics and problems identified in the underlying theoretical propositions.

3.1.1 The case study area. Bangalore city was selected as the case study area because it is one of the fastest growing economies in India, the opportunities the city offers as a result of this economic development has led to a significant growth in population. The infrastructure, transport and other amenities are insufficient to address this rapid growth and the city is witnessing congestion, high levels of pollution, deteriorating health, etc. Bangalore is a good representative of rapidly growing cities in developing countries.

3.2 Spatiotemporal analysis
It is well documented that geography plays an important role in understanding the dynamics of health and the causes and spread of disease and where a person lives determines their well-being. Geographical epidemiology is an effective tool to describe and analyse disease clustering correlated with demographic, environmental, behavioural, socio-economic, genetic and infectious risk factors. Based on this premise, this research proposes the use of a spatiotemporal methodology to assess the relationship between AP variables and mortality trends over a period of time. A GIS-based environment health information system is developed to identify the geographical distribution and variation in disease, correlate spatial and temporal trends of disease and AP and map populations at risk.

The proposed GIS environmental health system approach therefore integrates all of the temporal and spatial information to analyse and produce visualised tools, which facilitates greater communication (ALGA and ANZLIC, 2007). Consequently, this will allow policy makers to envisage the information described and thereby formulate more informed decisions. The proposed approach will incorporate the following:

- the development and use of an approach to enhance AP and CVD activities;
- allow various data and information resources to be integrated and analysed for the understanding of the influence of pollutants on CVD;
- make use of existing Big Data sets and use technology to assist in the awareness of CVD and the confounding factors;
• provide a graphical representation of the spread of pollutants and any CVD clusters;
• promote collaboration and learning;
• inform policy and advocacy;
• improve programmes, practice and research;
• enhance health training and education programmes;
• increase awareness; and
• propose interventions leading to an enhanced quality of life.

3.3 Data sets

Although the impact of the air problem in Bangalore is acknowledged by government officials, there is lack of an evidence base to highlight the city’s health scenario. The recent wealth of research that highlights the link between air pollutants and CVD is clearly absent in the city and Bangalore would certainly benefit from a research study that integrates data from the various departments.

The following section describes the different data sets that were collected as part of this research to provide a spatiotemporal analysis. This included the Spatial Data of Bangalore consisting of ward and zone maps. The demographic data of Bangalore from the BBMP was obtained that included 5,840,155 records from the 2001 census and 8,370,161 records from the 2011 census.

AP data were provided by the KSPCB for eight years for the pollutants SO₂, NOₓ and PM10 from six fixed locations.

The mortality data were obtained from BBMP. A total of 1m records (n = 1,090,899) was obtained. Of the n = 1,090,899 records, initial analysis determined that not all records had the cause of death recorded. The absence of cause of death does not assist in the analysis of death rates attributed to specific conditions such as CVD. Hence those records were deemed unuseful for the purposes of this research. The years 2010–2013 were selected for the purposes of the study which resulted in a data set with n = 183,893 as the previous years had no cause of death recorded. This data set was thus subjected to the assessment of data quality using the data quality parameters.

It was envisaged that a GIS-based system would enable the spatiotemporal analysis of AP and CVD, by integrating the data sets and graphically representing the results, the system also operates as a decision-support tool.

The data sets obtained for this research were Spatial Data, Demographic Data, CVD Mortality Data and AP Data.

3.3.1 Demographic data. The Bangalore Municipal Corporation known as Bruhat Bangalore Mahanagara Palike (BBMP) is the administrative body responsible for the civic and infrastructural assets of Bangalore city. The demographic data contained herein have been gathered from BBMP.

Furthermore, the 2011 census included a number of parameters such as growth rate in population, rate of literacy, density of population, sex ratio and child sex ratio (0–6 years). In Bangalore, the BBMP is responsible and is the administrative body responsible for the civic and infrastructural assets of the Bangalore metropolitan area. The population data were obtained from BBMP for the 2001 census and the 2011 census. The variables that were obtained were:

• name of the ward;
• ward number;
• zone;
Population in 2001. Data for the census of 2001 have been obtained from the BBMP for the 198 wards in Bangalore (Table I). The variables include total population, male and female population. The census data for 2011 for the same variables have also been obtained from BBMP (Table II).

Population in 2011. The population in Bangalore in 2001 was 5,840,155 of whom 3,059,580 were males and 2,780,575 females. By 2011, the population rose to 8,370,161 with 4,361,730 males and 4,007,755 females. This was an increase of about 2.5m people in just ten years contributing to 30.23 per cent growth in overall population. The male population contributed to 52.39 per cent of the total population in 2001 and female population 47.61 per cent.

Zone statistics. The 198 wards in Bangalore are divided into eight zones and this section discusses the zonal statistics for the population. Table III illustrates the population statistics for the eight zones in 2001 and 2011 and also provides the gender distribution of the population.

At zonal level, all zones witnessed an increase in population ranging from 10.02 to 61.25 per cent. Figure 1 and Table IV indicate that the zone population has witnessed explosive growth since the 2001 census. Zone Bommanahalli has the highest increase in population, at 61.25 per cent, followed by RR Nagar with an increase of 55.07 per cent.

<table>
<thead>
<tr>
<th>TOTAL_POP_2001</th>
<th>MALE_2001</th>
<th>FEMALE_2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>29,495.73</td>
<td>15,452.42</td>
</tr>
<tr>
<td>Range</td>
<td>17,295</td>
<td>10,627</td>
</tr>
<tr>
<td>Minimum</td>
<td>19,287</td>
<td>9,899</td>
</tr>
<tr>
<td>Maximum</td>
<td>36,582</td>
<td>20,526</td>
</tr>
<tr>
<td>Count</td>
<td>198</td>
<td>198</td>
</tr>
<tr>
<td>Sum</td>
<td>5,840,155</td>
<td>3,059,580</td>
</tr>
</tbody>
</table>

**Source:** BBMP

<table>
<thead>
<tr>
<th>TOT_POP_2011</th>
<th>MALE_2011</th>
<th>FEMALE_2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>42,273.54</td>
<td>22,028.94</td>
</tr>
<tr>
<td>Range</td>
<td>72,710</td>
<td>38,130</td>
</tr>
<tr>
<td>Minimum</td>
<td>21,120</td>
<td>10,721</td>
</tr>
<tr>
<td>Maximum</td>
<td>93,830</td>
<td>48,851</td>
</tr>
<tr>
<td>Count</td>
<td>198</td>
<td>198</td>
</tr>
<tr>
<td>Sum</td>
<td>8,370,161</td>
<td>4,361,730</td>
</tr>
</tbody>
</table>

**Source:** BBMP
Dasarahalli, Byatarayanapura and Mahadevapura followed with the highest population increases of 53, 51 and 46 per cent, respectively. The inner zones of the city on the other hand had lower population increases with the East, South and West having a 14.3, 19.23 and 10.225 per cent increase, respectively. The South zone of the city is the most densely populated with a population of 2,012m.

3.3.2 Air pollution data. AP data have been acquired from the KSPCB. The data for the years 2006–2013 have been obtained from the KSPCB as annual averages. The pollutants for which the data are consistently available are SO2, NOx and RSPM/PM10. The annual average values for these pollutants for all eight years were obtained for the six fixed locations in Bangalore:

1. Graphite India Limited, Whitefield Road;
2. KHB Industrial Area, Yelahanka;

<table>
<thead>
<tr>
<th>Zones</th>
<th>TOTAL_POP_2001</th>
<th>TOT_POP_2011</th>
<th>POP_INCREASE</th>
<th>INCREASE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bommanahalli</td>
<td>340,182</td>
<td>877,859</td>
<td>537,677</td>
<td>61.25</td>
</tr>
<tr>
<td>Byatarayanapura</td>
<td>348,464</td>
<td>725,794</td>
<td>377,330</td>
<td>51.99</td>
</tr>
<tr>
<td>Dasarahalli</td>
<td>265,012</td>
<td>575,529</td>
<td>310,517</td>
<td>53.95</td>
</tr>
<tr>
<td>East</td>
<td>1,208,688</td>
<td>1,410,359</td>
<td>201,661</td>
<td>14.30</td>
</tr>
<tr>
<td>Mahadevapura</td>
<td>374,195</td>
<td>698,295</td>
<td>324,100</td>
<td>46.41</td>
</tr>
<tr>
<td>RR Nagar</td>
<td>183,676</td>
<td>408,829</td>
<td>225,153</td>
<td>55.07</td>
</tr>
<tr>
<td>South</td>
<td>1,625,059</td>
<td>2,012,069</td>
<td>387,010</td>
<td>19.23</td>
</tr>
<tr>
<td>West</td>
<td>1,494,869</td>
<td>1,661,427</td>
<td>166,558</td>
<td>10.02</td>
</tr>
</tbody>
</table>

Source: BBMP
(3) Peenya Industrial area, Regional Office Peenya;
(4) Victoria Hospital, Chamrajpet;
(5) AMCO batteries, Mysore Road; and
(6) Yeshwanthpur Police Station.

Any values provided for mobile monitoring locations were not considered since obtaining a trend or association with intermittent values is not possible. The locations of the monitoring stations are entered as shape files and the values corresponding to each monitoring station are entered into an attribute file according to the years obtained. The variables used are as represented in Table V.

3.3.3 CVD data. The data register containing individual data are entered into a central database by the BBMP and the death registers are sent to the Department of Economics and Statistics which maintains a database used for statistical reporting. The CVD mortality data were obtained from the BBMP based on central repository data register that contains the mortality data integrated from all the registration centres in Bangalore. A total of 1,090,899 historic records were obtained for the years 1930–2013. Data were recorded according to year of death, zone, age, gender and cause of death. The details of the zone names and age groups are represented in Table VI. Data management tools (Excel and Access) were used for the management of the data prior to entering it into the ENVHIS.

The cause of death is coded according to the WHO's ICD-10 classification for diseases. Codes start from A00 to Z99, with CVD mortality codes ranging between I00 and I99.

3.4 Approach to data analysis

This research conducted spatial and aspatial analysis to analyse the effects of pollution on CVDs. However, this paper focuses predominantly on the spatial analytical techniques conducted on the ENVHIS. The aim was to provide an integration of data sets and provide visualisation on the data sets.

Data integration involves combining data residing in different sources and providing users with a unified view of this data. Using the ENVHIS system, data are combined and analysis performed to demonstrate the effects of AP on CVD in the study area. Hotspots of AP are determined, and relationships and patterns between environmental pollutant hazards and CVD are explored using the GIS. The spatiotemporal analysis provides an insight into how pollutants impact CVDs, and lead to the identification of areas and population groups that are most affected.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQM Station</td>
<td>Graphite KHB Victoria Peenya YPR AMCO</td>
</tr>
<tr>
<td>AQM station type</td>
<td>Industrial Residential Sensitive</td>
</tr>
<tr>
<td>Year</td>
<td>2010-2013</td>
</tr>
<tr>
<td>Pollutant name</td>
<td>SO₂ NOₓ PM₁₀</td>
</tr>
</tbody>
</table>

Table V. Variables for AQM
3.4.1 Brief discussion on air pollution analysis. Analysis on the AP data obtained for the eight years was carried out for the six locations for pollutants SO2, NOx and PM10 (Chinnaswamy et al., 2016). The analysis highlighted that the city is inundated with critical AP issues. While SO2 levels were well below the recommended standard, levels of NOx was either moderate or high; the main concern was PM10 that exceeded the recommended level in all areas of the city.

Most studies obtain recorded pollution levels from the monitoring stations and average them for the whole city. But the nature of pollution especially particulate matter is such that they are spatially variable, based on various factors, such as size of particle, wind, location of structures and other meteorological conditions. Many AP studies (Zhang et al., 2013; Matejícek et al., 2006; Son et al., 2010) have considered spatial interpolation methods to produce maps of AP concentrations. Interpolation is necessary when the data do not completely cover the domain of interest in order to predict the values of attributes at unsampled sites, by using the known measurements made at locations within the same area. The interpolated maps of SO2, NOx and PM10 and the subsequent inferences are discussed by (Chinnaswamy, 2015).

3.4.2 Brief discussion on CVD analysis. The statistical analysis of the data revealed that the total non-institutional deaths over the years 2010–2013 were 86,818 of which 33,075 deaths were due to CVD and 53,743 deaths were due to all other causes. CVDs over the four years on an average contributed to 38.10 per cent of all deaths. Table VII and Figure 2

<table>
<thead>
<tr>
<th>Zones</th>
<th>All deaths</th>
<th>Non-CVD</th>
<th>Per cent</th>
<th>CVD</th>
<th>Per cent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bommanahalli</td>
<td>2,873</td>
<td>1,641</td>
<td>57.12</td>
<td>1,232</td>
<td>42.88</td>
</tr>
<tr>
<td>Byatarayanpura</td>
<td>5,263</td>
<td>3,150</td>
<td>59.85</td>
<td>2,113</td>
<td>40.15</td>
</tr>
<tr>
<td>Dasarahalli</td>
<td>3,111</td>
<td>1,780</td>
<td>57.22</td>
<td>1,331</td>
<td>42.78</td>
</tr>
<tr>
<td>East</td>
<td>15,783</td>
<td>7,737</td>
<td>49.02</td>
<td>8,046</td>
<td>50.98</td>
</tr>
<tr>
<td>Mahadevapura</td>
<td>6,259</td>
<td>3,755</td>
<td>59.05</td>
<td>2,604</td>
<td>40.95</td>
</tr>
<tr>
<td>RR Nagar</td>
<td>4,572</td>
<td>3,168</td>
<td>69.29</td>
<td>1,404</td>
<td>30.71</td>
</tr>
<tr>
<td>South</td>
<td>23,428</td>
<td>15,998</td>
<td>68.29</td>
<td>7,430</td>
<td>31.71</td>
</tr>
<tr>
<td>West</td>
<td>25,429</td>
<td>16,514</td>
<td>64.94</td>
<td>8,915</td>
<td>35.06</td>
</tr>
<tr>
<td></td>
<td><strong>86,818</strong></td>
<td><strong>53,743</strong></td>
<td><strong>61.90</strong></td>
<td><strong>33,075</strong></td>
<td><strong>38.10</strong></td>
</tr>
</tbody>
</table>

Table VII. Non-institutional mortality by zones for years 2010–2013
represent the mortality distribution by geographic locations (i.e. zones). Zones West and East had the highest mortality followed by zones South and Mahadevapura. However, when the percentage of total deaths is considered, CVD deaths in zone East was the highest contributing to 50 per cent of the total deaths.

3.4.3 Rate of death. Mortality rates could provide an exact understanding of the frequency of deaths in a particular zone. This provides a general indication of the health status of a geographic area and assists stakeholders in healthcare planning for the population of that area. Hence, understanding the CVD mortality rate will enable the healthcare planning and allocating of resources and policies for an affected area in Bangalore.

This analysis highlighted that the zones West, East and South have the highest CVD mortality rates across the region Table VIII represents the analysis of deaths by age groups and gender. Although in absolute numbers the quantity of CVDs are highest for the age group 75+, when CVDs affecting early death, i.e., below 70 years of age the total number of deaths is 41,388 contributing to 66 per cent of total CVD deaths. This is an important factor to consider as the younger a person dies the more socio-economic impact there is on the individual’s family and the community at large.

Table IX provides the number of deaths by age group, gender and zone of death. This provides a comprehensive understanding of the number of deaths occurring at every zone by age group. As highlighted, the zones South, West and East have the highest number of CVD deaths with more males (62 per cent) being affected than females (38 per cent).

3.5 Spatial analysis
The previous section discusses a descriptive analysis of the extent of AP and CVD in Bangalore. However, integrating these data sets and a spatial analysis provides key stakeholders with visualisation to aid in decision making. Spatial analysis can be described as a set of techniques that aim to determine relationships between spatial and non-spatial variables. Spatial epidemiological analysis includes the determination of spatial patterns, identification of disease clusters and the explanation or prediction of disease risk (Pfeiffer, 2008).

<table>
<thead>
<tr>
<th>ZONES</th>
<th>BONMAHALI</th>
<th>HATAPRAKAPURU</th>
<th>DAWNAPALI</th>
<th>EAST</th>
<th>MAHADAVAPURA</th>
<th>RR NAGAR</th>
<th>SOUTH</th>
<th>WEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Of Deaths</td>
<td>16,000</td>
<td>14,000</td>
<td>12,000</td>
<td>10,000</td>
<td>8,000</td>
<td>6,000</td>
<td>4,000</td>
<td>2,000</td>
</tr>
<tr>
<td>NON-CVD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table VIII. CVD deaths by age groups and gender
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bommanahalli</td>
<td>156</td>
<td>24</td>
<td>42</td>
<td>40</td>
<td>59</td>
<td>89</td>
<td>142</td>
<td>147</td>
<td>207</td>
<td>230</td>
<td>271</td>
<td>257</td>
<td>575</td>
<td>2,239</td>
</tr>
<tr>
<td>F</td>
<td>64</td>
<td>8</td>
<td>15</td>
<td>9</td>
<td>17</td>
<td>17</td>
<td>37</td>
<td>41</td>
<td>62</td>
<td>76</td>
<td>102</td>
<td>74</td>
<td>269</td>
<td>791</td>
</tr>
<tr>
<td>M</td>
<td>92</td>
<td>16</td>
<td>27</td>
<td>31</td>
<td>42</td>
<td>72</td>
<td>105</td>
<td>106</td>
<td>145</td>
<td>154</td>
<td>154</td>
<td>183</td>
<td>306</td>
<td>1,448</td>
</tr>
<tr>
<td>Byatarayanpura</td>
<td>132</td>
<td>10</td>
<td>26</td>
<td>38</td>
<td>69</td>
<td>111</td>
<td>141</td>
<td>142</td>
<td>217</td>
<td>245</td>
<td>288</td>
<td>280</td>
<td>663</td>
<td>2,362</td>
</tr>
<tr>
<td>F</td>
<td>65</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>10</td>
<td>24</td>
<td>43</td>
<td>31</td>
<td>68</td>
<td>94</td>
<td>103</td>
<td>106</td>
<td>326</td>
<td>885</td>
</tr>
<tr>
<td>M</td>
<td>67</td>
<td>8</td>
<td>19</td>
<td>32</td>
<td>59</td>
<td>87</td>
<td>98</td>
<td>111</td>
<td>149</td>
<td>151</td>
<td>185</td>
<td>174</td>
<td>337</td>
<td>1,477</td>
</tr>
<tr>
<td>Dasarahalli</td>
<td>102</td>
<td>30</td>
<td>32</td>
<td>35</td>
<td>94</td>
<td>93</td>
<td>143</td>
<td>144</td>
<td>179</td>
<td>174</td>
<td>208</td>
<td>170</td>
<td>364</td>
<td>1,768</td>
</tr>
<tr>
<td>F</td>
<td>39</td>
<td>11</td>
<td>11</td>
<td>6</td>
<td>28</td>
<td>24</td>
<td>36</td>
<td>31</td>
<td>43</td>
<td>56</td>
<td>69</td>
<td>46</td>
<td>186</td>
<td>586</td>
</tr>
<tr>
<td>M</td>
<td>63</td>
<td>19</td>
<td>21</td>
<td>29</td>
<td>66</td>
<td>69</td>
<td>107</td>
<td>113</td>
<td>136</td>
<td>118</td>
<td>139</td>
<td>124</td>
<td>178</td>
<td>1,182</td>
</tr>
<tr>
<td>East</td>
<td>1,021</td>
<td>146</td>
<td>184</td>
<td>264</td>
<td>416</td>
<td>597</td>
<td>938</td>
<td>1,085</td>
<td>1,458</td>
<td>1,677</td>
<td>1,681</td>
<td>1,687</td>
<td>3,766</td>
<td>14,920</td>
</tr>
<tr>
<td>F</td>
<td>464</td>
<td>63</td>
<td>61</td>
<td>66</td>
<td>119</td>
<td>158</td>
<td>271</td>
<td>320</td>
<td>486</td>
<td>581</td>
<td>675</td>
<td>726</td>
<td>1,898</td>
<td>5,888</td>
</tr>
<tr>
<td>M</td>
<td>557</td>
<td>83</td>
<td>123</td>
<td>198</td>
<td>297</td>
<td>439</td>
<td>667</td>
<td>765</td>
<td>972</td>
<td>1,096</td>
<td>1,006</td>
<td>961</td>
<td>1,868</td>
<td>9,032</td>
</tr>
<tr>
<td>Mahadevapura</td>
<td>268</td>
<td>71</td>
<td>98</td>
<td>108</td>
<td>194</td>
<td>220</td>
<td>309</td>
<td>337</td>
<td>436</td>
<td>486</td>
<td>483</td>
<td>412</td>
<td>713</td>
<td>4,135</td>
</tr>
<tr>
<td>F</td>
<td>110</td>
<td>35</td>
<td>40</td>
<td>37</td>
<td>47</td>
<td>58</td>
<td>99</td>
<td>90</td>
<td>126</td>
<td>161</td>
<td>177</td>
<td>143</td>
<td>336</td>
<td>1,459</td>
</tr>
<tr>
<td>M</td>
<td>158</td>
<td>36</td>
<td>58</td>
<td>71</td>
<td>147</td>
<td>162</td>
<td>210</td>
<td>247</td>
<td>310</td>
<td>325</td>
<td>306</td>
<td>269</td>
<td>377</td>
<td>2,676</td>
</tr>
<tr>
<td>RR Nagar</td>
<td>137</td>
<td>39</td>
<td>41</td>
<td>46</td>
<td>92</td>
<td>112</td>
<td>162</td>
<td>170</td>
<td>222</td>
<td>244</td>
<td>267</td>
<td>218</td>
<td>511</td>
<td>2,261</td>
</tr>
<tr>
<td>F</td>
<td>52</td>
<td>19</td>
<td>14</td>
<td>11</td>
<td>18</td>
<td>37</td>
<td>47</td>
<td>54</td>
<td>60</td>
<td>72</td>
<td>93</td>
<td>70</td>
<td>229</td>
<td>776</td>
</tr>
<tr>
<td>M</td>
<td>85</td>
<td>20</td>
<td>27</td>
<td>35</td>
<td>74</td>
<td>75</td>
<td>115</td>
<td>116</td>
<td>162</td>
<td>172</td>
<td>174</td>
<td>148</td>
<td>282</td>
<td>1,485</td>
</tr>
<tr>
<td>South</td>
<td>1,677</td>
<td>342</td>
<td>443</td>
<td>476</td>
<td>748</td>
<td>938</td>
<td>1,325</td>
<td>1,492</td>
<td>1,827</td>
<td>2,087</td>
<td>2,316</td>
<td>1,993</td>
<td>4,264</td>
<td>19,928</td>
</tr>
<tr>
<td>F</td>
<td>730</td>
<td>146</td>
<td>174</td>
<td>141</td>
<td>238</td>
<td>274</td>
<td>400</td>
<td>463</td>
<td>619</td>
<td>747</td>
<td>930</td>
<td>803</td>
<td>2,157</td>
<td>7,822</td>
</tr>
<tr>
<td>M</td>
<td>947</td>
<td>196</td>
<td>269</td>
<td>335</td>
<td>510</td>
<td>664</td>
<td>925</td>
<td>1,029</td>
<td>1,208</td>
<td>1,340</td>
<td>1,386</td>
<td>1,190</td>
<td>2,107</td>
<td>12,106</td>
</tr>
<tr>
<td>West</td>
<td>1,056</td>
<td>196</td>
<td>266</td>
<td>287</td>
<td>501</td>
<td>708</td>
<td>957</td>
<td>1,134</td>
<td>1,335</td>
<td>1,501</td>
<td>1,707</td>
<td>1,663</td>
<td>3,993</td>
<td>15,304</td>
</tr>
<tr>
<td>F</td>
<td>452</td>
<td>83</td>
<td>90</td>
<td>72</td>
<td>124</td>
<td>183</td>
<td>270</td>
<td>326</td>
<td>412</td>
<td>505</td>
<td>693</td>
<td>652</td>
<td>2,060</td>
<td>5,922</td>
</tr>
<tr>
<td>M</td>
<td>604</td>
<td>113</td>
<td>176</td>
<td>215</td>
<td>377</td>
<td>525</td>
<td>687</td>
<td>808</td>
<td>923</td>
<td>996</td>
<td>1,014</td>
<td>1,011</td>
<td>1,933</td>
<td>9,382</td>
</tr>
<tr>
<td>Grand total</td>
<td>4,549</td>
<td>858</td>
<td>1,132</td>
<td>1,294</td>
<td>2,173</td>
<td>2,868</td>
<td>4,117</td>
<td>4,651</td>
<td>5,881</td>
<td>6,644</td>
<td>7,221</td>
<td>6,680</td>
<td>14,849</td>
<td>62,917</td>
</tr>
</tbody>
</table>
Heywood (2006) states that the spatial analysis techniques in GIS provide concepts and transformations in such a way that is not possible with other techniques. A vast range of spatial analysis techniques can be carried out on the data that may include simple queries or more complex geospatial analyses. The results of the analyses can be viewed on the database table or in the form of charts, graphs or reports. Examples of spatial analysis techniques include:

- buffer analysis;
- overlay analysis;
- geocoding;
- spatial interpolation;
- spatial regression; and
- hot spot analysis.

The spatial analysis was carried out on the data in the ENVHIS using the techniques described in the following sub-sections.

3.5.1 Spatial data. The base maps (major roads, water bodies and railway lines) were digitised using 1:50,000 topographic sheets obtained from Survey of India. Survey of India is a National Survey and Mapping Organisation responsible for mapping the country’s domain and providing base maps (Survey of India 2012). The ward maps were obtained from BBMP in the .jpg format and later converted to .dwg format and finally to .shp format (UTM Projection).

The coordinates of the air quality monitoring stations were captured using a Garmin hand-held GPS and plotted onto the map with the same projection (UTM) system. The thematic maps consisting of base maps, ward delimiters, zone delimiters, highways and major road networks, water bodies and rail network were entered into the ENVHIS as .shp files consisting of wards and zones as polygons; roads and rail networks as lines; air monitoring stations as points. The attributes for these mainly include the name, area, length, latitude, longitude, etc.

The variables for demographic data, AP data and CVD data were coded in the system and the values entered respectively. Using the GIS integration techniques, the integrated data were presented as layers, allowing for analysis and visualisation. The spatial analysis included the following:

3.5.2 Geocoding. Geocoding is a method of classifying geographic location. It involves the process of assigning $x$ and $y$ coordinates based on the actual latitude and longitude of the earth’s surface where the location is sited (Summerhayes et al., 2006). Geocoding technology often resides within the more complex GISs that manage location-based and related data.

In the ENVHIS addresses were converted to features on the map through the geocoding process (Figure 3). The steps involve translating the address entry, searching for the address in the reference data and delivering it as a feature on the map.

There are different types of geocoding process such as point entry, street address, etc. In this geocoding process, due to the nature and availability of the data, the location assigned to a particular address is the polygon representing the geographic unit. Location within the unit is not specified but analysis is carried out using the data associated with the geographic unit.

3.5.3 Dot density maps. Once the geocoding process is complete, the dot density maps are created. Dot density maps or dot maps represent the geographic distribution of discrete phenomena using point symbols, most commonly identical dots. These maps particularly provide an understanding of the distribution of the phenomena under consideration by
Figure 3.
Geocoding by polygon
comparing relative densities of different regions on the map. There are two kinds of dot density maps:

1. one-to-one maps; and
2. one-to-many maps.

In the one-to-one dot density maps each point on the map corresponds to a single incidence of the mapped phenomena. One-to-many dot density maps, on the other hand, represent a pre-determined number of incidences of the mapped phenomena. To represent the CVD deaths, the one-to-many dot density maps are used as the data have been aggregated to an enumeration unit and there are many points within the units for representation.

The four important design considerations for one-to-many dot density maps are:

1. units of aggregation;
2. dot size;
3. dot values; and
4. dot placement.

The unit of aggregation is at zonal level (Figure 4). The dot size and value are chosen carefully: small dots would produce an overly sparse dot pattern. Dots that are too large would produce excessively dense dot patterns.

3.5.4 Overlays. Overlay is a spatial operation in which two or more maps or layers are superimposed for the purpose of showing the relationships between features that occupy the same geographic space. Figure 5 illustrates an overlay of PM$_{10}$ and CVD deaths for the year 2010. Visually, the highest numbers of CVD deaths are concentrated in the middle of the city.
3.5.5 Queries. In order to answer questions such as “What areas have PM$_{10}$ higher than the standards and high number of CVD deaths?”, the query can be designed in ENVHIS using SQL expressions. The results of these queries are then displayed both on the map and the table. Building a query expression is a way to select features as an expression can include multiple attributes, operators and calculations. The queries can be saved as separate layers. Figures 6 and 7 provide results of a query both in a table and on the map.

3.5.6 Reports. A report in GIS allows organising and displaying the tabular data associated with geographic features. This can be generated and distributed along with the map. Using the report wizard in ArcMap, the fields from the layer necessary to be presented as a report are chosen. All the fields to display can be chosen or the selected fields or even a subset of data based on a query may be chosen for reporting. The properties of the report may be chosen by the user too, such as style of the report, page size, font or colour.

3.5.7 Heatmaps. Heat mapping, from a geographic perspective, is a method of showing the geographic clustering of a phenomenon. Also known as hotspot mapping, heat maps show locations of higher densities of geographic entities. The “heat” in the term refers to the concentration of the geographic entity within any given spot, and is not to be confused with heat mapping that refers to the mapping of actual temperatures on the earth’s surface. Heat mapping is a way of geographically visualising locations so that patterns of higher than average occurrence of things such as crime activity, traffic accidents, or store locations can emerge.

3.5.8 CVD + PM$_{10}$ hotspots. Due to the data limitations as mentioned earlier, the hotspots were discerned at the zonal level. The zones West and East were found to be hotspots with the highest CVD deaths and the highest PM$_{10}$ deaths. However, the system would be more effective if the ward hotspots are identified so as to target interventions at those levels. Improvements in data collection methods and data quality would assist in focussing primary and secondary interventions in places and populations that are at most risk.
Figure 6.
Tabulated queries
Figure 7.
Query on map and table
4. Discussion

In this paper we discuss the growing magnitude of the health problem in Bangalore as a result of air pollution induced CVDs. Hence, an ENVHIS is developed to integrate the Big Data sets from various sources that will bring together geospatial knowledge and information under a common platform, which till date has existed as a disparate system and unknown to quite a wider section of the society, institutions, scientific community and government departments. The ENVHIS also enables information sharing across the various data generating agencies, government departments, NGOs, academies, industries and scientific organisations. This will then facilitate strategic decision making and help in local level planning.

The ENVHIS system assists in the identification of hotspots of AP which then enables the government to take actions to reduce AP in the zones at most risk first; this can be achieved through the:

- development of realistic policies and regulations for air emissions from transportation sources;
- use of ultra-low sulphur fuels;
- introduction of regular and reliable public transportation systems;
- improvement of the existing mass transit system;
- robust spot-checking, inspection and mandatory testing programmes for vehicles;
- stricter control on road-side burning of waste; and
- better maintenance of roads to reduce dust from paved and unpaved roads.

The ENVHIS system also assists in the identification of hotspots of CVD which then enables the government to introduce interventions in the zones at most risk first, and this can be achieved through:

- development, implementation and monitoring policies that reduce the prevalence of CVD;
- reduce inequalities in risks of developing CVD;
- offer comprehensive advice or appropriate treatment to reduce risks;
- identify populations who have not yet developed CVD but are at risk to offer appropriate advice to prevent it;
- increased awareness among public health officials, doctors, health workers on the possible links of AP and CVD, and encourage to include it is a risk factor and offer advice on related health improvement; and
- help raise public awareness about the health impacts of AP.

It is any government’s priority to provide public health programmes to face the challenges of environmental-related health issues. The WHO (2011) states that over 50 per cent of the CVDs can be prevented with early interventions. Quality treatment, early diagnosis, good quality medical and nursing care, specialist services – such as surgery and rehabilitation for those with CVD or who are at risk – will facilitate patient recovery and/or well-being. The government as a major priority must focus on prevention, diagnosis and treatment.

Policies must evolve to incorporate the conclusions of important new research as it becomes available. Preventive approaches require a combined effort. Pais (2006) states that risk factors tend to cluster together and it is important to identify risk factors for CVD. Co-existing risk factors multiply the risk of disease, e.g., tobacco smoking resulted in an OR for Myocardial Infarction of 3.1. The OR tripled in diabetics who smoked and tripled again if
people additionally suffered from hypertension. Further investigations will determine the OR (risk factors) in the case of adding pollutants to this model.

The Indian government must initiate legislative and economic policies to increase taxes on tobacco and saturated fat, subsidise fruit and vegetables and make available facilities for walking, cycling and exercise. This requires strong advocacy and political will.

Pandey et al. (2013) state that health-related poor lifestyles are widely prevalent in low- and middle-income countries such as India. Prevalence of smoking and tobacco is high, intake of dietary saturated fats; trans-fat and salt are high while fruits and vegetables intake is low. Behavioural factors such as smoking/tobacco use, physical activity, unhealthy diets lead to metabolic changes such as hypertension, raised cholesterol, raised blood glucose and obesity (WHO, 2011b).

There have not been significant population-wide efforts to influence healthier lifestyles. Interventions among populations are important as better knowledge of dietary factors involved in chronic disease can lead to transformational influence on diet and other lifestyle behaviours that can have a direct impact on these diseases in the society and also be economically justifiable. For interventions to take effect it is important to make the interventions people-centric.

5. Conclusions
This paper has demonstrated a methodological approach for the collection and visual representation of Big Data sets allows for an understanding of the spread of CVDs across the city of Bangalore, enabling different stakeholders to query the data sets and reveal specific statistics of key hotspots where action is required. Awareness campaigns on signs of CVD, knowledge of treatment options, nearest healthcare centres that deal with response to CVD would enable tackling and reduction of fatalities from the disease.

The developed GIS-based ENVHIS enables the understanding of the correlations between CVD and mortality. CVD has a large impact on the economy and society at large. The system aids in identification of areas at high risk which then enables the government to tackle the issue by understanding the cost of CVDs for economy and society. The biggest challenge in understanding the magnitude of the problem is the absence of data in the right formats. This research advocates on improve data collection and related processes to inform future research (aid strengthening both the study of AP and health management). Improvements in the data collection process and highlighting the objectivity of the data collection process would enable not only further research, but also result in other advantages such as better health management and economic impact for the government of Bangalore.

The spatial data analysis has provided a visual correlation of AP to CVD mortality and highlighting the hotspots for both pollution and mortality. This enables informing future healthcare management strategies and contributes significantly in formulating future healthcare management strategies and health policy.

Awareness campaigns on factors affecting CVD, knowledge of treatment options, nearest healthcare centres that deal with response to CVD would enable tackling and reduction of fatalities from the disease. Interventions can be targeted in areas of dire need rather than rolling out to the whole city which may both be expensive and impractical.

Health interventions are termed as a success only when they have provided a beneficial change in the health of the population. Where there is no beneficial change, the appropriateness of the intervention provided requires investigation. Hence, it is vital to monitor the outcome of health interventions in order to not only determine the change in health levels of the population but also to measure the quality and effectiveness of the healthcare provision. The impact can be measured as economical, societal or individual. The economic impact measures how the interventions and the subsequent results had an impact on the economy. Ill health adversely affects the development of human capital, which is crucial for a developing economy. The societal impact focuses on the creation, sharing and application of the knowledge derived from
the growing volumes of data potentially available to for the benefit of society. Many of the risk factors that contribute to CVD are further compounded by underlying socio-economic factors as well as environment factors. The right investments will lead not only to better health, but also to longer and more productive lives. This will impact the individual, the individual’s family, society and community at large.

References


Further reading


About the authors

Anitha Chinnaswamy is Senior Lecturer in the Coventry Business School and Research Associate of the Centre for Business in Society (CBiS). With a background in Computing and Information systems (PhD), and Health Informatics (MSc), she has been involved in research and consultancy projects with collaborators within the UK and international partners. Her area of research pertains to data and information management relating to social computing that include resilience and preparedness, environmental policy and management, cyber security strategy and management. Anitha Chinnaswamy is the corresponding author and can be contacted at: anitha.chinnaswamy@coventry.ac.uk

Armando Papa is Research Fellow at University of Turin. He holds PhD in Management achieved at the University of Naples Federico II. He holds a Postgraduate Master in Finance (2011). He is also Member of the EuroMed Research Business Institute. He won the Best Paper Award at the 9th EuroMed Conference in Warsaw (Poland) in 2016. He is skilled in Knowledge Management, Entrepreneurship and Corporate Governance and Family Business Management. He is Associate Editor of Journal of Knowledge Economy Springer and Editorial Assistant of Journal of Knowledge Management Emerald. Engaged in various peer-review process for several ranked and outstanding international management. He is member of the I.P.E. Business School of Naples.

Luca Dezi is Full Professor of Economics and Business Management at the University “Parthenope”, Faculty of Economics, Naples. Furthermore, he is Professor of Strategic Marketing and International Business at the Faculty of Communication Sciences, Department C.O.R.I.S., University of Studies of Rome, Sapienza. Member of the Company’s Board of Auditors San Giovanni Addolorata Hospital in Rome from 2007 to 2011. He is also an Ordinary member of the Italian Academy of Business Economics - Chartered accountant and Auditor. He is a Consultant in the field of Economic and Financial Analysis, Planning and Management Control, Evaluation of Businesses and Projects, Corporate Restructuring, Mediations, Rationalisation and Resource Management, Teaching and Training on behalf of medium-sized and large companies, public institutions and organisations.

Alberto Mattiacci is Full Professor of Management and Marketing at the Department of Communication and Social Research (Coris), University of Rome “Sapienza”. He is Co-Director of Master Marketing Management (www.mumm.it) and Director of Master Digital Export Manager at Agenzia ICE (Italian Trade Agency). He is President of the Italian Marketing Society (www.simktg.it) after having served as General Secretary from 2011. He is engaged in numerous Scientific Board of private and public research centres. He is also Scientific Director at Eurispes, a private centre for research on change in Economics, Politics and Society and co-founder and member of Circle International, an academic network on distribution and consumer studies and member of the advisory board of the Institute for Business, Government and Culture. He is visiting at Villanova Business School at Philadelphia.

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
The integration between knowledge management and dynamic capabilities in agile organizations

Fábio Lotti Oliva, Marcelo Henrique Gomes Couto and Ricardo Fernandes Santos
Department of Business Administration, Universidade de Sao Paulo, Sao Paulo, Brazil, and
Stefano Bresciani
Department of Business Administration, Universita degli Studi di Torino Dipartimento di Management, Torino, Italy

Abstract
Purpose – The purpose of this paper is to describe and analyze how the integration between knowledge management and dynamic capabilities in contexts that demand organizational agility contributes to the management by objectives.

Design/methodology/approach – To achieve the proposed objective, the present paper adopts the single case study in the startup Effecti. For data collection, semi-structured interviews were carried out, analyzed a posteriori through the technique of content analysis. From the loads of evidence observed, a model was presented that consists of different management theories and that guides the management by objectives process of a startup.

Findings – The proposed model proves to be able to describe the modus operandi of a startup and enables it to develop the cycles of testing, measurement and seizure of knowledge, largely stimulated and inherent to the creation process of new businesses in dynamic and uncertain contexts.

Practical implications – It is expected that the research results presented in details can illustrate concrete examples of application of the main concepts: agile organization, dynamic capabilities, knowledge management, performance assessment, enterprise risk management and management by objectives.

Originality/value – The originality of this study is focused on the integration of conceptual triad and its application in the case study of a startup: agile organization, dynamic capabilities and knowledge management.

Keywords Dynamic capabilities, Knowledge management, Enterprise risk management, Agile organization, Performance assessment, Management by objectives

Paper type Research paper

1. Introduction
All organizations exist for a purpose and, to achieve this, senior management plans and sets common goals and objectives from the vision of its founders and executives. Startups, for example, are nascent, implementation and organizational institutions that are set up to seek, test and build a business model that is repeatable, scalable and profitable, but that operate in highly uncertain ecosystems from diverse perspectives and present limited resources and rapid growth which are vulnerable to a number of sources of risk, and therefore have high mortality rates (Blank and Dorf, 2014; Giardino et al., 2014; Haarigan et al., 2012; Ries, 2012; Picken, 2016).

Achieving the goals in a startup’s current business environment can be a difficult task, however, especially given the current level of uncertainty and rapid transformations.
So how do we manage and achieve the goals of these institutions in their current uncertain and rapid development contexts?

Thus, due to the risks acquired since its initial conception, the conditions of uncertainty and rapid evolution they are subjected to, new untapped challenges are presented to the startuppers (Blank, 2013; Giardino et al., 2014; Ries, 2012), requiring rapid transformations in their business models as a response to pressures and tensions caused by internal and external processes.

Thus, although in the business environment and history it is common for managers to be challenged to produce results in a context of risks and uncertainties, in addition to delivering them, the management must also be concerned to produce them under a larger context of uncertainty and rapid technological and social changes and to use these rapid changes in favor of their businesses, minimizing transaction costs and maximizing transaction value (Felipe et al., 2016; Schotter et al., 2017; Teece et al., 2016; Thomson, 1998; Hwang and Gaur, 2009).

From this perspective, therefore, in order to achieve its objectives successfully, a startup needs to develop individual and organizational learning capacities during the process of business development (Sosna et al., 2010) and management models that seek to integrate different management approaches that meet the specificities of managing their context and minimize the chances of failure.

Thus, given the lack of market mechanisms that protect entrepreneurial innovation from the uncertainties associated with the reception of an opportunity (Janeway, 2012), agility seems to be an important attribute or organizational capability to be developed by current organizations seeking sustainable development (Baskarada and Koronios, 2018; Felipe et al., 2016; Ries, 2012; Singh et al., 2013; Teece et al., 2016), considering the increase in the organization’s ability to proactively respond to unexpected environmental changes, developed by agility (Appelbaum et al., 2017).

Therefore, the companies that have achieved success are those that, in addition to meeting latent demands at the right time, are fast and flexible enough in terms of innovation (Teece et al., 1997), supported by the foreknowledge capability (prescience), and by the tools used to find out which parts of the business plans are misguided, in order to adapt the strategies accordingly (Ries, 2012). And it is in this context that some recently popular and promising methods have emerged and promised to increase agility and reduce risks in the decision-making process and to achieve the business objectives in uncertain environments (Blank, 2013; Blank and Dorf, 2014; Ries, 2012; Teece et al., 2016).

Thus, the agility required and inherent to these methods contributes and makes easy the search and the retrieval of relevant knowledge, allowing companies to rapidly apply such knowledge to develop high-quality solutions or react to the emergence of new competitors (Cegarra-Navarro et al., 2016). However, despite the innumerable success stories associated with agile development, a number of barriers, ranging from technical to perceptual barriers related to the process of implementing agile development, reveal the importance of the management commitment in the process of providing the necessary objectives and support aiming at the successful development of such a process (Boehm and Turner, 2003; Pikkarainen et al., 2012) as well as the need for a tailor-made, risk-oriented approach so that projects take advantage of the planned benefits of agile methods while mitigating some of their disadvantages (Boehm and Turner, 2003).

Being so, strong dynamic capabilities are critical for a company to promote organizational agility required to operate in an environment of uncertainty, since they are the ones that define the company’s ability to innovate, adapt and create changes that are favorable to the consumer market and unfavorable to competitors (Teece et al., 2016).

However, the development of dynamic capabilities depends heavily on the organization’s knowledge and learning (Easterby-Smith and Prieto, 2008), which, in turn, is strongly
related to the role that key individuals play in the process of knowledge development within an organization (Nuruzzaman, Gaur and Sambharya, 2018). For this reason, knowledge management must be a practice developed in organizational processes and widely stimulated by governance, in order to take its effectiveness and facilitate its transfer (Anderson et al., 2015), thus improving organizational processes continuously. In this way, the knowledge management expands the competitive capacities of the organizations (Anderson et al., 2015; Nuruzzaman, Gaur and Sambharya, 2018; Nuruzzaman, Singh and Pattnaik, 2018) and contributes to the achievement of the organizational objectives or results and value creation potential (Mukherjee et al., 2018; Mukherjee, Gaur and Datta, 2013), also being measured as an organizational performance measure (Oliva, 2014; Qi and Chau, 2018).

Despite these developments and the set of pragmatic methods based on lean and agile principles recently proposed to support entrepreneurs in the evolution and innovation of their business models, little has been investigated on the theoretical and practical relationship between these management practices and the evolution and innovation of these business in dynamic environments, especially under the support and integration of some traditional and recent approaches to business administration (Ghezzi and Cavallo, 2018; Oliva et al., 2011).

From this perspective, the objective of the present paper is to describe and analyze how the integration between knowledge management and dynamic capabilities in contexts of agile organizational development contributes to the management by objectives.

2. Theoretical framework

2.1 Management by objectives

The entrepreneur must understand quite clearly his objectives in order to keep an organization focused on them (Picken, 2016). Thus, initially developed by Peter Drucker in 1954, the approach known as management by objectives has been used in different ways since then, in order to assist management team in the process of defining and orienting the specific objectives to be achieved. Besides Drucker, this approach has also been developed by many other management theorists, including Douglas McGregor, George Odiorne and John Humble, in which, in essence, who enshrine the process designed to jointly set specific goals to be accomplished within a defined deadline and followed by the definition of the responsibilities aiming at reaching the agreed targets and analyzing the process performance (Campbell, 2015; Thomson, 1998).

Thus, the approach has been widely used as a management system that aligns tangible objectives with the organization’s vision (Dinesh and Palmer, 1998), based on a set of strategic planning parameters that seeks to harmonize management objectives with the employees’ objectives (Islami et al., 2018) and use the objectives achievement as a guide to assess the performance of the employee and of the organization as a whole (Campbell, 2015).

However, for the management by objectives approach achieves satisfactory results of productivity and performance, it is necessary that the objectives setting, the participation in decision making and the objective feedback combine with the top management’s approach and commitment to it. Top management commitment has to be high, since proper implementation starts at the top and requires top management support and participation (Rodgers and Hunter, 1991). Therefore, not by chance, management by objectives has been translated as a process (Islami et al., 2018; Thomson, 1998) that requires coherence of goals through collaboration (Dinesh and Palmer, 1998).

Thus, as the market acceptance of management principles has been broadened, a number of stages have been developed for its implementation, allowing the approach robust application (Dinesh and Palmer, 1998). These authors also present a six-stage structure that begins with the strategy identification and the long-term objectives, development of action
plans and allocation of resources up to the performance constant review aiming at assuring the results.

2.2 Agile organization
Wide disseminated in the context of software development, the principles announced by the movement known as agile manifesto encourage practices that allow changes in requirements at any stage of the solution development process, and actively engage customers or users in that process, facilitating the feedback and iterativity, which consequently may lead to more satisfactory results (Dingsoyr et al., 2012).

However, despite the principles that translate themselves as guidelines for the development of high-quality and agile solutions (Dingsoyr et al., 2012), a more robust definition based on a comprehensive literature review is presented by Singh et al. (2013). The authors define the concept as systematic and persistent variations in an organization’s products, structures or processes that are identified, planned and executed as a deliberate strategy to gain competitive advantage.

Therefore, in the organizational context, one can understand by agility the ability to respond flexibly to changes in the environment, adjusting quickly the offerings of products and services (Felipe et al., 2016; Henderson-Sellers and Serour, 2005; Singh et al., 2013) as well as redirecting resources, in an efficient and effective way, in order to create, capture and protect value in higher income activities (Teece et al., 2016). Therefore, in this sense, organizational agility acts as an organizational capability that facilitates the integration and organization of resources and knowledge and not only the rapid application of knowledge (Cegarra-Navarro et al., 2016).

Thus, from this philosophy and based on the principles it communicates, different methods for creating and developing projects and businesses have become popular in the organizational context, as a way of developing products or services in an iterative and incremental way (Blank and Dorf, 2014). Thus, because they deal with the unpredictability and with the focus on human potential, these methods are characterized by short and iterative cycles in which there are planning, development, testing and deliveries, and such cycles are powered by the addition of incremental improvements, periods of reflection and introspection, rapid feedback incorporation and focus on changes when needed (Campanelli and Parreiras, 2015; Nerur et al., 2005).

In addition, there should be emphasized that since the commitment to continuous transformation and agile strategies imply changes at all organizational levels, a number of factors such as structure, strategy, capabilities, employees and leadership can also, in many ways, affect the organizational agility (Appelbaum et al., 2017). This, consequently, may require different organizational configurations leading, for example, a complex mix of organizational structures and leadership styles (Baskarada and Koronios, 2018).

In a broader context, global organizations are steadily struggling to capture the changing business environment from their various fronts. Transaction costs arising from contracting a new global supplier, outsourcing production and services to markets that have the most competitive comparative advantage, comply with new legislation in the markets for which it exports, research and development of products to meet the demands sites with the proper balance with meeting the demands of global standards, upgrading their systems due to new local or global technology, etc., are some examples by which global organizations are submitted. An organization that wants to remain competitive in this global scenario should minimize transaction costs and generate more value in its local and global transactions. In this sense, the development of relationships with local and global organizations is an important strategy to meet these goals. In this way, global organizations must be agile and have the dynamic capabilities to compete in a volatile, uncertain, complex and ambiguous global environment (Mukherjee et al., 2018; Schotter et al., 2017;

Therefore, agile development involves the characteristics of flexibility, speed, learning and response to changes (Blank and Dorf, 2014; Campanelli and Parreiras, 2015; Giardino et al., 2014).

2.3 Dynamic capabilities
Originally proposed by Teece et al. (1997), the understanding of the concept found in the literature, however, is not universal (Ambrosini and Bowman, 2009; Schilke et al., 2018; Wang and Ahmed, 2007). Featuring some inconsistencies, similarities, differences and conceptual contradictions (Barreto, 2010), it is observed that sometimes the researchers direct the emphasis to understanding the concept as an organizational ability (Teece et al., 1997), and sometimes such researchers direct the emphasis to understanding the concept as an organizational routine (Eisenhardt and Martin, 2000), or as a set of processes (Ambrosini and Bowman, 2009) or as a behavioral orientation of the company (Wang and Ahmed, 2007).

In spite of the different perspectives for the subject, however, it is common as the general understanding that the dynamic capabilities result from the combination and reconfiguration of the management processes developed by the organization over time through a learning routine that is a result of the organizational trajectory and that allows the organization to perceive the environmental changes, in order to adapt its activities according to the established needs (Eisenhardt and Martin, 2000; Teece et al., 1997). Thus, dynamic capabilities are the managerial capabilities needed to sustain and adjust the organizational resources basis toward the organizational efficiency achievement (Eisenhardt and Martin, 2000; Teece, 2007; Teece et al., 1997) and govern as an integrated company, and build and reconfigure internal and external competencies (Teece et al., 2016).

In this sense, strong dynamic capabilities are critical for a company to promote the organizational agility required to operate in an environment of uncertainty, since these uncertainties are the ones that define the company’s ability to innovate, adapt and create changes that are favorable to the consumer market and unfavorable to competitors (Teece et al., 2016; Nuruzzaman, Singh and Pattanaik, 2018). This invariably involves management’s ability to perceive and then seize opportunities and manage threats by combining and recombining assets to sustain the development of long-term value (Schilke et al., 2018; Teece, 2007).

Thus, the dynamic capabilities are the company’s potential to solve problems systematically, from the perception of opportunities and threats, in order to make timely and market-oriented decisions (Barreto, 2010) and to ensure competitive advantage and longevity to the company (Eisenhardt and Martin, 2000; Kumar et al., 2018; Teece et al., 1997, 2016). Therefore, from this perspective, dynamic capabilities can be a source of competitiveness in face of market demands (Ambrosini and Bowman, 2009; Teece et al., 1997), which reflect the integration of individuals’ expertise in the organization (Adegbite et al., 2018), culture, orientation and leadership (Schilke et al., 2018), and the redefinition and reorganization of corporate strategies in order to meet identified needs (Kumar et al., 2018).

2.4 Knowledge management
Knowledge is the most important resource for innovative organizations (Papa et al., 2018). Depending on the organization’s objectives, it can be used to develop different forms of value creation and results, such as refining, renewing, recombining and replicating a strategy, and, therefore, its management is a practice developed in organizational processes to bring about its effectiveness and to create further value in dynamic environment (Mukherjee et al., 2018; Mukherjee, Gaur, Gaur and Schmid, 2013; Mukherjee, Gaur and Datta, 2013).
Through the knowledge formalization, therefore, organizational processes are continuously improved. Thus, the knowledge management expands the organizations’ competitive capabilities and contributes to the achievement of the organizational objectives, also being considered as an organizational performance measure (Oliva, 2014; Qi and Chau, 2018).

Thus, the lack of knowledge management in an organization can lead to the learning degeneration and, consequently, to a reduction in the effectiveness of organizational processes (Oliva, 2014).

According to Probst et al. (2002), knowledge management is performed through a schematic process comprising eight key steps to be developed:

- definition of knowledge objectives;
- knowledge identification;
- knowledge acquisition;
- knowledge development;
- knowledge dissemination;
- knowledge use;
- knowledge retention; and
- knowledge assessment.

However, the performance of these steps can be compromised by barriers of several natures. The main barriers are of human nature and are linked to human and organizational behavior in the internal context. They occur mainly in the acquisition, dissemination and assessment of knowledge (Oliva, 2014).

Thus, as the practices of knowledge management are implemented by companies, they can be characterized in a certain stage of the knowledge management maturity (Oliva, 2014). The author presents four stages of maturity for knowledge management:

1. The initial level is characterized by low adherence to the practices related to knowledge management.

2. The structured level deals with organizations with some concern about the importance of knowledge management for the business, in which there is a greater use of knowledge management practices, supported by the existing information systems. At this level, the collaborative participation is still low.

3. The oriented level not only refers to a profile of companies equal to the previous level, but also presents an orientation for the creation and use of knowledge to boost the business, in addition to an organizational culture oriented toward innovation.

4. The last level, integrative, indicates high concern, organization and transparency regarding knowledge management. There is the use of outside consultants and other external support aiming at improvements in knowledge management. Finally, knowledge management goes beyond organizational boundaries, considering the knowledge dispersed in the intersecting environment of organizations.

2.5 Performance assessment

Performance assessment in new product development involves a number of observations including: the company’s innovation culture, its strategy, its portfolio management, its development tools, its metrics and its results (Markham and Lee, 2013). In this sense, conventional models of the product development process allow an assessment based on performance and results (Cooper, 2014).
However, uncertainty is a matter of course in the context of startups and establishing an assessment process is something complex that requires more appropriate observations about the organizational potential and performance (Ries, 2012). An assessment process should therefore consider strategic and team issues as well as financial issues such as capital invested, cash flow, net current value and sensitivity to scenarios. As far as the startup strategy is concerned, the following must be considered: the value supply, the size of the explored market, the sustainability of the competitive advantages and the legal issues that imply the business. As for the professional team's analysis, it considers characteristics such as: experience and motivation of the founders, availability of capital and contribution of investors. This more holistic analysis allows not only to assess a specific project, but also to understand the entrepreneurship applied to the product development process in a startup (Csaszar et al., 2006).

2.6 Enterprise risk management
In the current global business context, organizations are immersed in an environment with a degree of uncertainty related to the external environment market, strategy and planning issues (Schotter et al., 2017). The solutions to these issues are mostly sized for large corporations, and because they present a different profile, they are not applicable to the reality of startups (Ries, 2012). In addition, there is the fact that professionals who participate since the beginning, from the ideation to the development of technological solutions are paid a low compensation (Hall and Woodward, 2010). Thus, startups naturally are more likely to take risks and, consequently, higher failure rates (Giardino et al., 2014).

Enterprise risk management (ERM) plays a facilitating role in preventing negative outcomes, thus becoming important an accurate phase of identifying risks and threats to the organization and its projects (Stosic et al., 2016; Singh and Delios, 2017). Some sectors are most exposed to risks and increased volatility, and this can cause proportional impacts to organizations, especially in agile development environments (Hall and Hulett, 2002). Thus, risk management becomes critical to generating gains in these environments (Teberga et al., 2018).

To guide the implementation of ERM in an organization, Oliva (2016) has structured the Maturity Model in Risk Management, in which he took into consideration the following items for determining the maturity level: the processes of planning, organizing, implementing, risk management, technicity and transparency of these processes, as well as the organization’s involvement in ERM performance analysis and support. Thus, the ERM maturity level practiced in organizations can be classified as:

- insufficient: with little awareness about risk management;
- contingency: companies are already aware of the risk management they face and present a modest use of ERM tools and methods;
- structured: it presents a greater degree of organization of the processes related to ERM;
- participatory: high level of awareness and some organization of ERM processes; and
- systemic: ERM conscious organized and transparent.

2.7 Integration of approaches
Managing a business requires the entrepreneur to be clear about their goals and current situation in order to develop a clear work program and direction to keep an organization focused on the appropriate goals (Picken, 2016). In other words, an administration by objectives is applied. In startups, however, given the conditions and characteristics
associated with it, as well as the objectives of rapid growth (Blank and Dorf, 2014; Giardino et al., 2014; Ries, 2012), organizational agility to reach goals and innovate demonstrated a valuable attribute due to the constant state of transformation (Ghezzi and Cavallo, 2018; Teece et al., 2016).

However, in order for the organization to achieve the required levels of agility, it must develop strong dynamic capabilities to enable it to cope with deep uncertainty and promote the ability to respond rapidly to market changes (Teece et al., 2016). Thus, given the context of uncertainties and rapid environmental transformations in which startups are embedded, it becomes fundamental that collective knowledge be adapted and current enough to cope with these environmental conditions, or the company’s chances of survival diminish (Teece et al., 2016; Sosna et al., 2010).

Therefore, a knowledge management mechanism that supports this process and promotes the ability to perceive and respond rapidly to changes, as well as transform and adjust strategies (Kumar et al., 2018), resources and parts of the business model (Blank and Dorf, 2014; Sosna et al., 2010; Teece et al., 2016) becomes a necessity for management, as well as a control system that measures and evaluates the changes that are addressed to the development of a sustainable business and the institution’s value delivery objectives (Ries, 2012).

In this way, as risks are associated with known results and objectives (Teece et al., 2016) and startups naturally present a greater tendency to take risks and, consequently, higher failure rates (Giardino et al., 2014), the development of a risk management model may play a facilitating role in preventing negative outcomes (Stosic et al., 2016), and thus complement the set of management approaches required to develop a startup.

3. Methodology
3.1 Research design
This research presents a single case study (Yin, 2014) of a qualitative nature (Cooper and Schindler, 2003; Miles et al., 2014), which seeks to understand in detail the phenomena approached, its regularities and exceptionalities, through an interview with five executives of Effecti. The research also has an observatory-experimental character, since it integrates theoretical concepts related to organizational management with a focus on agile organizations.

3.2 Conceptual model
Startups are temporary and early-stage development institutions. They are created and organized to search, test and build a business model that is repeatable, scalable and profitable, from the initial vision of its founders (Blank and Dorf, 2014; Haarigan et al., 2012; Ries, 2012). However, in order to achieve this, the entrepreneur needs to be clear about his objectives and to keep the institution focused on them (Picken, 2016), since, in addition to exploiting new opportunities under a highly uncertain market environment (Giardino et al., 2014; Ries, 2012), startups are still characterized by limited resources and personnel, third-party dependency and time pressure (Giardino et al., 2014) (component 1 – management by objectives).

Therefore, in addition to a management system that can align tangible goals with the organization’s vision (Dinesh and Palmer, 1998), startups must develop the capacity to react to environmental changes, grow and expand rapidly (Giardino et al., 2014) in order to meet the rapid transformation needs of its business models as a response to the pressures and tensions caused by internal and external processes (component 2 – agile organization).

In this respect, strong dynamic capabilities are thus key to enabling startup to rapidly develop, adapt and renew its business model in order to create value on a sustained basis
(Teece et al., 2016) (component 3 – dynamic capabilities). Thus, since this process results from the combination and reconfiguration of the management processes developed over time, through a learning routine that is a consequence of the organizational trajectory of a company and that enables it to perceive environmental changes (Eisenhardt and Martin, 2000; Teece et al., 1997; Zollo and Winter, 2002), it is also important to integrate a knowledge management mechanism to support this process (component 4 – knowledge management).

However, even if these processes are taken care of by the startup, so that it ensures that previously planned objectives are achieved and, consequently, the chances of failure decrease, it is also fundamental that the institution develops a holistic analysis that allows it to evaluate its project and its level of maturity and to understand the entrepreneurial capacity applied to the startup development process (Csaszar et al., 2006) (component 5 – performance assessment). And in this regard, ERM can play a facilitating role in these businesses in preventing negative outcomes, since it allows identifying risks and threats to the organization’s success (Stosic et al., 2016; Teberga et al., 2018), contributing to better management of the conditions and characteristics of a startup (component 6 – enterprise risk).

Given the context presented, the phenomenon addressed in this research, as presented in Figure 1, addresses the processes of management by objectives, agile organization, dynamic capabilities, knowledge management, performance assessment and enterprise risk management performed by Effecti.

3.3 Data collection

For data collection, two interviews were conducted, the first of which was an open and personal script with the founder and CEO of Effecti and the second, with a semi-structured script through video conference, with the same founder and CEO and four other executives of Effecti. The interviews, according to Cooper and Schindler (2003) and Yin (2014), involved

![Figure 1. Research conceptual model](source: Prepared by the authors)
issues based on the literature, presented in Figure 1, which guided the dialogue between interviewer and interviewee. Aspects such as neutrality, monitoring of the interviewee’s line of reasoning and request for details were considered during interviews.

The semi-structured script of the interview was based on the literature review in management by objectives, agile organization, dynamic capabilities, knowledge management, performance assessment and ERM.

During the interview, they were asked about the processes related to the topics discussed in the literature review. The five executives interviewed were selected, in accordance with Cooper and Schindler (2003), and according to their experiences and unique perceptions of the phenomenon analyzed. The interviews took place through a videoconference with an average length of 200 min. Limiting the number of interviews was due to the facts confirmation. The main methodological limitation in this study was the impossibility of carrying out personal interviews.

3.4 Presentation and data analysis
The scope in which the proposal of this paper is presented seeks to integrate management approaches and, from this, to advance in the understanding of a theoretical management model applied to the context of startups. Thus, as the objective advances the frontiers of current knowledge in management, the technique chosen for data analysis was content analysis; therefore, it allows and applies to studies and interpretations of qualitative data to develop themes from studies of single case (Gaur and Kumar, 2018).

Thus, data analysis was based on the methodology of Miles et al. (2014). After being collected, the data were coded, according to the research conceptual model, structured from the literature review proposed (Gaur and Kumar, 2018; Miles et al., 2014), and jointly allocated allowing to compare their regularities and exceptionalities. With the data allocated and codified, the analysis was performed with the internal triangulation of data. Figure 2 schematically shows the developed content analysis process.

4. Analysis of results
4.1 Description of the case
Effecti is a startup that provides technology solutions to bidders. In other words, it provides digital solutions that connect supplier companies to the government. Found in September 2013, this startup currently serves more than 1,100 companies in all Brazilian states.

Data collection:
Personal interview with CEO
Videoconference interview with CEO and more four executives

Data coding:
Definition of the coding scheme based on the topics of literature review and coding of all transcribed texts from interviews

Interpretation of coded content:
Comparison, internal triangulation and interpretation of coded data

Source: Prepared by the authors based on Gaur and Kumar (2018) and Miles et al. (2014)
In addition, of the 100 largest solution providers to the Brazilian Federal Government, 37 percent use Effecti solutions in their operations.

The startup is based in the city of Rio Sul, state of Santa Catarina, and in São Paulo, state of São Paulo, and it is one of the residents of Banco Bradesco co-innovation space, the Habitat. Furthermore, Effecti has recently participated in the innovative business acceleration program carried out by the Brazilian Ministry of Industry, Commerce and Services (MDIC) and by the Brazilian Service for Support to Micro and Small Enterprises (SEBRAE), the Inovativa Brasil.

The business initial idea arose from past professional experiences of the founders in the field and the opportunity to develop a solution for a company that supplied products to the government. These experiences, added to the opportunity to apply an idea to a potential customer, enabled them to identify routine difficulties of the supplier companies associated with the excess of bureaucracy related to this process and the possibility of developing solutions that automate and reduce the time spent in performing tasks during the bidding process.

In order to verify the market viability, the solution was presented to other potential customers who, in part, contracted the technology. The huge and rapid growth of the solutions, however, is recent, and is due to, in particular to the increase of the technical and commercial team and the technological resources improvement of the solution needed to meet a growing demand. Effecti has a current staff of 14 people.

Currently, the startup offers four different systems to assist the phases of the bidding process. The development of these distinct solutions was a consequence of the focus on meeting other suppliers’ needs and of the ability to identify other repetitive routines that could be automated. Moreover, these solutions must meet at least one of two principles: increase profits or reduce the time to complete the tasks and, hence, reduce costs. They are offered through free trial versions for a specified period of time and can be purchased by subscription models.

4.2 Management by objectives
During the startup development process, the entrepreneur must be clear about his objectives and his current situation in order to establish and communicate a clear direction to keep the organization focused on the appropriate objectives (Picken, 2016). Thus, the management strategies in Effecti are always developed as from the main needs inherent to the company’s development at a certain moment of its evolution and are delineated based on the resources, capabilities, objectives and main problems to be solved in the current moment. Currently, for example, the company’s main objectives are related to the process of the vertiginous and rapid growth of its solutions and seek to solve the company’s growth problems, since the startup is in this stage.

Therefore, the development process of the startup set of objectives starts from the identification of the company’s current strategy, followed by the preparation of action plans and objectives, the provision of resources and assignments of responsibilities (Dinesh and Palmer, 1998). For Effecti, in this process, a clear understanding of the company’s current capabilities and resources is also important, given the limitations of a startup and the need to align tangible objectives with the organization’s vision (Dinesh and Palmer, 1998).

In short, the objectives not only guide the company’s activities but also direct the focus, the main resources and the development of solutions internally and external to the organization.

4.3 Agile organization
An agile organization is one that identifies, plans and executes rapid changes in its solutions, structures or processes (Singh et al., 2013) as a response to the constant
environmental changes (Felipe et al., 2016; Henderson et al., 2005; Singh et al., 2013) and addition of incremental improvements, periods of reflection and introspection, rapid feedback incorporation and focus on changes when needed (Campanelli and Parreiras, 2015; Nerur et al., 2005). Thus, seeking to adapt the need to develop a greater capacity to respond to the required changes in a faster way, one of the processes management methods in development is an adaptation of Scrum.

Thus, some assumptions of the method are applied and some are not. This option for the development and adoption of an adaptation of the method is an evidence of response to the need for startup in learning how to measure team performance and maximize the fulfillment of the company’s objectives.

In addition, organizational flexibility is an important feature to stimulate agility (Campanelli and Parreiras, 2015; Giardino et al., 2014; Teece et al., 2016). For this, the startup adopts a more horizontal and organic structure with industry leaderships and more flexible systems that facilitate the apprehension and rapid adaptation in face of the changes. However, so that the speed is preserved, it is essential that people are committed, the areas must be developed collaboratively and the organizational objectives, besides being evident, must be able to guide all the company’s activities.

4.4 Dynamic capabilities

Uncertainty, with regard to business sustainability, is an inherent feature of Effecti’s environment. Despite having its own technology, the solution requires the operation of government platforms and the consumption of information arising from third-party databases, which are beyond the company’s domain and that can compromise the operation. Moreover, solutions can be heavily influenced by changes in legislation, which highlights one of the business vulnerabilities.

Thus, seeking to reduce this vulnerability and promote the business sustainability (Eisenhardt and Martin, 2000; Picken, 2016; Teece et al., 1997, 2016), the startup has diversified with regard to the development of new solutions and projects in areas complementary to its initial proposal, in addition to seeking to approach the key partners in its operation environment, such as government agencies, for example.

These loads of evidence, in principle, demonstrate the importance of developing capabilities that enable the organization to realize and then take advantage of opportunities and manage threats by combining and recombining assets to sustain the long-term value development (Kumar et al., 2018; Teece, 2007) through the development of an entrepreneurial management (Teece et al., 2016) and through the ability to identify and explore emerging opportunities and new sources of competitive advantages (Kumar et al., 2018; Schilke et al., 2018).

4.5 Knowledge management

Knowledge management occurs when the organization structures a process where steps such as the definition of knowledge objectives, the identification, acquisition, dissemination, use, storage and knowledge assessment are performed (Probst et al., 2002). Effecti neither presents this structure in a formalized way, nor does it perform all stages of the knowledge management process. However, some practices are performed informally and not systematized in order to meet the needs for accessing the knowledge whenever necessary.

This lack of formalization of the knowledge management process occurs mainly because of the organic nature and the lean startup structure, which allows the exchange of information and knowledge in an easier way (Teece et al., 2016). Thus, even without the formalization of knowledge management processes, Effecti managers are concerned with aspects that involve knowledge of the company (Papa et al., 2018), and some practices for knowledge formation are performed such as: brainstorms and meetings to identify the
know-how required to offer new solutions and improvement of the current ones and creation of own methods of management and dissemination, for example.

Institutionalization of knowledge management, however, is in process of formation, in response to the needs inherent to the rapid growth process and to the needs of information storage. In addition, the increase in the number of employees is an aspect aimed at making the knowledge management process formalized. Corroborating with Oliva (2014), the interviewees agree that from this point of the company’s development, the non-formalization of knowledge management will have negative consequences such as deterioration of learning and decrease of the effectiveness of organizational processes.

These loads of evidence suggest that Effecti presents a need to overcome the structured degree of knowledge management and move to the advanced degree presented by Oliva (2014). They also corroborate the points made by Picken (2016), when suggesting that the process of large and rapid growth of a startup requires structures, processes and discipline to sustain the accelerated growth.

4.6 **Performance assessment**
Effecti’s performance is assessed by metrics related to the degree of task execution, adhesion of new user to testing, retention, and conversion of test users into customers. Effecti’s management uses few metrics for performance assessment but such metrics reveal process effectiveness, employee performance and customer satisfaction, going against what Ries (2012) calls accounting for innovation and actionable metrics for development and learning measurement.

Furthermore, metrics for the performance assessment of employees and of the company as a whole are monitored daily and discussed at regular meetings, where cases of underperformance are addressed. It is highlighted the agility and integration of the managers team for monitoring and decision making in the performance assessment. Thus, Effecti’s performance assessment is carried out in a holistic way, seeking the effectiveness of the company as a whole (Csaszar et al., 2006).

4.7 **Enterprise risk management**
Given the nature of the uncertainties in Effecti’s business environment, one may notice that these uncertainties can strongly influence the agility of its organizational processes and impact on results (Giardino et al., 2014). So, given its uncertain environment related to the public sector, the startup has developed a system of mapping agents and risk, monitoring the main risks involved in its business, especially those related to public management, that have greater impact on its strategies and solutions offered to corporate customers (Teberga et al., 2018).

Less comprehensive risks, such as those associated with customer data security and loss of market share for new competitors, are mitigated when the company seeks to minimize the possibility of such occurrence, with concrete actions for data security, market monitoring and improvement of the solutions offered.

Thus, as the startup moves toward high growth objectives, more attention is given to the maturation of more proactive risk management processes, since fast-growing companies are particularly vulnerable to many sources of risk (Picken, 2016). Thus, ERM provides direct support in the identification, assessment and monitoring of risks so that organizational objectives are not impaired and, consequently, the delivery of company value is not compromised (Oliva, 2016; Oliva et al., 2014).

Therefore, in general, Effecti presents a structured ERM, according to Oliva (2016), highlighting the considerable degree of awareness about the risks related to the business and the considerable structuring of risk management processes.
4.8 Integration of approaches

Due to the dynamic environment in which startups are embedded, it is observed that the constant planning of activities required during the business evolution needs to be fast enough to keep up with the organizational agility required to operate, innovate and adapt to the constant changes required by these businesses context (Teece et al., 2016), as well as achieving the objectives related to its development.

Therefore, strong dynamic capabilities are fundamental to this process (Teece et al., 2016), since they enable the company’s ability to configure and reconfigure its basis of organizational resources toward organizational efficiency (Eisenhardt and Martin, 2000), Teece et al. (1997) and toward the solution of the operational and managerial challenges inherent to the startup maturity process (Picken, 2016).

Thus, the dynamic capabilities are added to the agile development to play an important role in the process of not only quick planning required for these businesses, but also of organization and allocation of organizational resources in order to eliminate waste and create value (Ries, 2012). In other words, this integration contributes to the encouragement of experimentation process largely stimulated by emerging agile methods (Blank and Dorf, 2014; Ries, 2012; Teece et al., 2016), enabling the startup to not only learn with the obtained results but also with the process itself and how to run it best.

Thus, the learning process can be facilitated by the agility, since agility contributes in a way that the knowledge search and recovery process can enable the company to adjust (Cegarra-Navarro et al., 2016) and regroup itself (Ries, 2012; Teece et al., 2016) in a fast way, so as to meet not only the demands required by the external environment, but also the organization’s own needs in its evolution. However, to do so, this validated learning needs to be managed.

So knowledge management, although empirical and not completely structured, can support the execution of these processes in the startup, since from its practice the organizational processes are continually improved, and its competitive capabilities are expanded, allowing that its organizational objectives are achieved, as well as its effectiveness (Anderson et al., 2015; Nuruzzaman, Gaur and Sambharya, 2018; Nuruzzaman, Singh and Pattanaik, 2018; Oliva, 2014; Qi and Chau, 2018). However, learning can only be validated if the startup develops assessment measures that allow it to understand its entrepreneurial capability applied to the development process of its solutions (Csaszar et al., 2006) and to measure accurately its current position and which advances were obtained after running an experimental process (Ries, 2012).

Furthermore, in order to move toward its objectives, a startup still needs to develop a proactive risk management system, since fast-growing companies are particularly vulnerable to risk sources (Picken, 2016). Thus, ERM provides support to the startup control process in order to reduce the risks that could prevent the startup from achieving its organizational objectives and, consequently, compromise the delivery of business value.

These loads of evidence, therefore, draw up the company’s behavior and show how goal-oriented management of a startup can be adapted to the dynamic context of startups and strengthened by the agility in the planning process by the dynamic capabilities in the organization process, by the knowledge management in the execution process and by the performance assessment processes and ERM in the control process (Figure 3). This model is also able to demonstrate in theory how the cycle developed by Eric Ries (2012), known as build-measure-learn, is managerially structured in startups.

5. Conclusions

5.1 Attendance the research objectives

In general, as well as in corporations, the organizational objectives of a startup vary in the course of its evolution. This variation describes the business orientation to the solution of
the main operational and management challenges inherent to a particular time of the business development. Thus, the objectives of a startup not only make up its organizational trajectory but also guide the activities and processes to be developed by it, and therefore, the way the company is conducted, the measure that shapes its activities and priorities, as well as directs the necessary resources and efforts.

In addition, the development of a startup is an experimental and iterative process, and therefore, during its history, the startup empirically develops its own routines and organizational processes, according to the learning it acquires from its mistakes and successes and the established objectives. So, therefore, the learning generated by a startup is a set of knowledge and capabilities that enable the business to reach its established objectives with the agility required in the dynamic context in which it is inserted, reducing risks and preserving scarce resources.

In this sense, the objective of describing and analyzing how the integration between knowledge management and dynamic capacities in contexts of agile organizational development contributes to the management by objectives was care, by presenting from the empirical evidences the way in which the management approaches proposed by the study are integrated.

### 5.2 Attendance to methodological procedures

The methodological procedures were met by integrating the qualitative approach of the study to the single case study method and the content analysis technique for data analysis.

In addition, the semi-structured script of the interview was based on the literature review in management by objectives, agile organization, dynamic capabilities, knowledge management, performance assessment and ERM, and the five participating executives were selected according to their experiences and perceptions of the phenomenon analyzed.
5.3 Contributions to the theory
The main theoretical contribution of the study is the presentation of a theoretical framework of startup management that integrates different management approaches necessary for business development characterized by scarce resources, fast transformation environments and uncertainties, and accelerated growth.

In this sense, therefore, the study advances the proposed theoretical limits by integrating the approaches of management by objectives, agile organization, dynamic capabilities, knowledge management, performance assessment and ERM, contributing to the understanding of a theoretical management model applied to the context of startups.

5.4 Management implications
From these observed loads of evidence, an applicable management theoretical framework was presented which consists of several management approaches and that guides the management by objectives process of a startup. So, when used in an integrated manner, this model is able to describe the startup modus operandi and enable it to develop the cycle of experimentation, measurement and knowledge apprehension, widely stimulated and inherent in the process of creating new business in dynamic and uncertain contexts. In other words, it contributes to a better understanding of the management model that meets the conditions and development needs of a startup, minimizing risks and chances of failure.

5.5 Limitations and future study opportunities
Future studies that seek to investigate the application of the theoretical model proposed in other startup cases in order to promote comparisons and validation are welcome. The use of a single case is a limitation of this study, but the method was justified by the search for the depth of results. Other case studies in different business environments at different times may contribute to the improvement of the proposed integration model between knowledge management and dynamic capabilities in agile organizations. In addition, quantitative approach studies that identify the main variables of each construct of the theoretical framework will also be fundamental for the solidification of the model.

References


About the authors

Fábio Lotti Oliva holds Bachelor Degree in Computer Science from University of São Paulo, and a Masters and PhD Degree in Business Administration from University of São Paulo. He is currently Associate Professor at the School of Economics, Business and Accounting at the University of São Paulo. He is Visiting Professor at the Université Pierre-Mendes-France and the Shanghai University. His main publications include: Journal of Manufacturing Technology Management, Journal of Knowledge Management, International Journal of Production Economics, Journal of Organizational Change Management, Management Decision and Benchmarking: An International Journal, and so on.

Marcelo Henrique Gomes Couto holds Master’s Degree in the Postgraduate Program in Administration of the Faculty of Economics, Administration and Accounting of the University of São Paulo (FEA-USP) and Participant in the Research Group on Enterprise Rich Management (FEA-USP). He holds a Bachelor’s Degree in Biological Sciences from the Federal University of Lavras (UFLA-MG) and from the Federal Institute of Education, Science and Technology of Minas Gerais (IFMG-Bambuí). He also holds a Postgraduate Degree in Marketing Management from the SENAC University Center (SENAC-SP). His academic experiences are related to research areas in entrepreneurship, innovation, risk management and strategy.

Ricardo Fernandes Santos has Master’s Degree in the Postgraduate Program in Administration of the Faculty of Economics, Administration and Accounting of the University of São Paulo (PPGA/FEA/USP).
with CNPq scholarship. He participates in the Group of Research in Enterprise Rich Management (FEA/USP) where he develops research on the subject. He graduated in Business Administration from the Federal University of Grande Dourados (UFGD), and has scientific articles published in national congresses and magazines. He was a CNPq grant holder in the Local Agents of Innovation Program (SEBRAE & CNPq), in which he worked with small business innovation advisory and research.

Stefano Bresciani, PhD in Business Administration, is Full Professor in Business Management at the School of Management and Economics, University of Torino, where he teaches Innovation management, and he is the Coordinator of PhD program in Business & Management. He undertakes research integrated with the Department of Management of the University of Torino; his main areas of research include innovation management, international business and open innovation. He is Vice President of the EuroMed Research Business Institute (EMRBI) and Chairman of the EMRBI Research Group on “Multinational enterprises and corporate governance”. He has published in many refereed journal articles, contributed chapters and books and presented papers to conferences on a global basis.

Knowledge management and dynamic capabilities 1979
Management Decision
Vol. 57 No. 8, 2019
pp. 1980-1992
© Emerald Publishing Limited
DOI 10.1108/MD-07-2018-0829

Big data for business management in the retail industry

Gabriele Santoro
Dipartimento di Management, Universita degli Studi di Torino, Turin, Italy
Fabio Fiano
Link Campus University, Rome, Italy
Bernardo Bertoldi
Dipartimento di Management, Universita degli Studi di Torino, Turin, Italy, and
Francesco Ciampi
Universita di Firenze, Florence, Italy

Abstract
Purpose – The purpose of this paper is to shed light on how big data deployment transforms organizational practices, thereby generating potential benefits, in a specific industry: retail.
Design/methodology/approach – To achieve the paper’s goal, the authors have conducted several semi-structured interviews with marketing managers of four retailers in Italy, and researched secondary data to get a broader picture of big data deployment in the organizations.
Findings – Data analysis helped identify specific aspects related to big data deployment, data gathering methods, required competences and data sharing approaches.
Originality/value – Despite the growing interest in big data in various fields of research, there are still few empirical studies on big data deployment in organizations in the management field, and even fewer on specific sectors. This research provides evidence of specific areas of analysis concerning big data in the retail industry.
Keywords Big data, Retail industry, Data and knowledge
Paper type Case study

Introduction
The competition in certain industries corroborates the need for firms to develop new processes to achieve success and sustainability. Businesses in the new century have been permeated by an increased digitalization of processes, and firms not following this phenomenon often lose competitive advantages (Nambisan et al., 2017). This digitalization has focused attention upon the importance of data and knowledge within organizations (Dennis et al., 2001), suggesting that leveraging high-quality large data and talent analytics throughout value chains are necessary to develop business strategies and increase performance (Gaur et al., 2011; Gaur et al., 2014; Scuotto, Santoro, Bresciani and Del Giudice, 2017). However, to accomplish the task of being a digital organization, firms need to change the decision-making culture and cultivate abilities, skills and competences within the organization (Brynjolfsson and McAfee, 2012; Sivarajah et al., 2017). Big data is considered to be the “next big thing in innovation” (Gobble, 2013), as it has potential for business value creation. Big data can create actionable ideas for delivering sustained value by providing smart data and establishing competitive advantages (Wamba et al., 2015) as it allows for the enhancement of data-driven decision making and the processes of organizing, learning and innovating at various levels (Wamba et al., 2017). Despite the growing interest in the adoption of digital technologies as well as Internet of Things (IoT) systems in several fields of research ranging from engineering and informatics to general management (Del Giudice, 2016; Del Giudice et al., 2016; Santoro, Vrontis, Thrassou, and Dezi, 2018), a clear picture of the dynamics of application and benefits of big data deployment is missing in the literature, especially in the business and management field. In particular, empirical studies on this topic are scarce. Research in this field mostly consists of recently published conceptual studies (Wamba et al., 2015; Wamba and Mishra, 2017). A few exceptions exist but
mostly regarding manufacturing firms (Merchant and Gaur, 2008; Acharya et al., 2018), the health industry (Wang, Kung and Byrd, 2018; Wang, Kung, Wang and Cegielski, 2018) and the food industry (O'Connor and Kelly, 2017). Consequently, more and more service firms struggle to understand how to deploy big data analytics and how to extract value from raw data.

As a result, this research aims to explore how big data deployment transforms organizational practices, thereby generating benefits, in a specific industry: retail. To really understand the benefits of big data, it is necessary to examine the managerial, economic and strategic impacts of big data analytics and explore the effective path of how big data analytics can be leveraged to deliver business value. It was decided to focus on a specific industry because big data deployment should be considered and analyzed according to an industry’s peculiarities to fully get the benefits and dynamics of the competitive and strategic environment. This is a major contribution of this study, and the hope is that future studies address the issue on the same level, in other industries.

To reach the paper’s goal, an empirical qualitative analysis through multiple case studies has been developed. Specifically, several semi-structured interviews were conducted with marketing managers of four retailers in Italy. The findings of this research highlight some advantages and managerial challenges of big data deployment to further the knowledge on this topic and to provide implications for managers. The retail industry was chosen as the context of analysis for its peculiarities and high level of competitiveness and, in particular, because data are considered essential in this industry (www.forbes.com/forbes/welcome/?toURL=https://www.forbes.com/sites/bernardmarr/2015/11/10/big-data-a-game-changer-in-the-retail-sector/&refURL=https://www.google.it/).

The paper is structured as follows: the next section provides a theoretical background on the key hallmarks of big data deployment. Then, the empirical study is described, explaining the research method and presenting the findings drawn from the semi-structured interviews with marketing managers. The final section provides conclusions and discusses the implications and limitations of the research.

Theoretical background
Technological development in the past decades significantly increased the volume of available data for organizations, having an impact on management decisions (Ciampi, 2008; Ciampi, 2017a, b; Kane et al., 2015). As a consequence, scholars have placed attention on how firms implement new ICTs to create value (Davenport, 1993; Lopez-Nicolas and Soto-Acosta, 2010; Bresciani et al., 2017; Scuotto, Santoro, Papa and Carayannis, 2017). This is due to the increasing importance of data, information and knowledge for a firm’s competitiveness (Boisot and Canals, 2004; Del Giudice and Straub, 2011; Del Giudice and Della Peruta, 2016; Sumbal et al., 2017). In particular, big data has achieved increasing attention. Big data consists of data sets whose size is beyond the capacity of typical database software tools to capture, store, manage and analyze (Ohlhorst, 2012; Weinberg et al., 2013). Big data is capable of both changing competition by “transforming processes, altering corporate ecosystems, and facilitating innovation” (Brown et al., 2011, p. 2) and unlocking organization business value by unleashing new organizational capabilities and value (Davenport et al., 2012), and by facilitating firms to tackle their key business challenges (Kiron et al., 2014).

The main attributes related to the concept of big data are volume, speed and variety (McAfee et al., 2012), as organizations must be able to face large volumes of data, especially a diverse range of data and sources, quickly and effectively (Brynjolfsson and McAfee, 2012; Ohlhorst, 2012; El-Kassar and Singh, 2018).

Wamba et al. (2015) define big data as a holistic approach to manage, process and analyze five Vs (i.e. volume, variety, velocity, veracity and value) in order to create actionable insights for sustained value delivery, measuring performance and establishing competitive advantages. Thus, the authors add the concepts of veracity, the importance of
quality data and the level of trust in various data sources, and value, the importance of extracting economic benefits from data.

Big data deployment can be aligned with the business intelligence tools that are required to provide intelligent aid for organizational processes. The great volume of data necessary must be acquired, filtered, stored and analyzed to become heterogeneous. Because the processes of filtering and analyzing the data are very complex, it is necessary to use business intelligence strategies and tools. In detail, big data management is defined as “a collection of data and technology that accesses, integrates, and reports all available data by filtering, correlating, and reporting insights not attainable with past data technologies” (APICS, 2012) and is emerging as a strategic aspect for firms (Brynjolfsson and McAfee, 2012; Alfouzan, 2015). More specifically, the concept of big data has been used to describe data sets that are so complex they cannot be managed or analyzed using traditional data analysis software (Waller and Fawcett, 2013), making it necessary for firms to outsource big data infrastructure and hire skilled employees (Santoro and Usai, 2018). In fact, because of the abundant information that such data sets contain, big data usually consists of large, heterogeneous and unstructured data sets that cannot be explored by the usual approaches (Yin and Kaynak, 2015). Compared with traditional databases, big data includes a large amount of unstructured data that must be analyzed in real time. Big data also brings new opportunities for the discovery of new values that are temporarily hidden (Renu et al., 2013).

Big data management and analytics allow managers “to measure and know radically more about their business and to directly translate that knowledge into decision making and performance” (Brynjolfsson and McAfee, 2012, p. 4). Managers may take decisions on the basis of big data analytics, which can shape decision making and business processes (Provost and Fawcett, 2013).

Acquiring and understanding data allow firms to extrapolate information that must be analyzed to develop new knowledge that can be used in management decisions (Dezi et al., 2018), such as customization, pricing options and the opening of new point of sales. There are two main features of data (Azma and Mostafapour, 2012). Smart processing includes analyzing and assessing the information, providing decision support to ensure that the future performance of the organization is aligned with the planning, as well as knowledge feedback about the processes involved to be combined with pre-existing (explicit) knowledge. The organizational learning includes the discovery of new knowledge and the dissemination of this knowledge to those who need it.

Overall, it has been suggested that big data deployment can improve internal processes (Davenport, 2014), as well as enhance operational agility and flexibility (Lu and Ramamurthy, 2011; Gao, 2013; Wamba and Mishra, 2017). Moreover, acquiring and transforming raw data in smart data can actually provide firms with vital knowledge about customer behavior (Hofacker et al., 2016; Motamarri et al., 2017).

Empirical research on big data confirms its direct effect on firm performance (e.g. Akter et al., 2016; Gupta and George, 2016; El-Kassar and Singh, 2018). However, capabilities and IT resources alone may not unequivocally foster firm performance (Melville et al., 2004), therefore calling for deeper and more exploratory research methods, especially in specific sectors.

**Research design**

*Research context: retail industry*

Retailing is the group of activities performed when offering or selling products or services among consumers for their personal, family or institutional use (Kotni, 2011). Specifically, a retail firm is an organization that aims to acquire the product or service in demand to deliver it later to the consumer (Pantano, 2014). Such industry represents the foundation and health of a formal or informal economy within a region or country, giving final consumers access to products and services that otherwise would be difficult to get (Kotni, 2011).
The retail industry in Italy contributes to creating employment all over the country, and despite the economic and financial crisis which has led to the worst financial result in 2014 in this sector, the overall turnover has been increasing in the last three years, led especially by food commerce (www.dgmco.it/media/studies/pdf/gdo-report-2016-2017.pdf).

The retail industry is considered well suited for this research for its peculiarities and high level of competitiveness and, in particular, because data are considered essential in this industry. For example, gathering large amounts of data about clients is seen essential to improve the competitiveness of retail firms due to the customized product offering that data can drive and customized prices that can be applied.

It was specifically decided to focus on the retail industry (Saghir and Jönson, 2001). The last years have brought profound changes in this industry, with customers increasingly careful about what they buy and consume, about prices and so on. As clients become more aware and informed about products, prices and trends, the value chain of retailers become more complex (Peker et al., 2017). Moreover, this industry is more and more threatened by low margins and fierce competition, pushing toward new digital processes to increase efficiency and adaptation. Data deployment help in managing such complexity.

Methods of analysis
The epistemological foundation of this study is based on the interpretivist approach. The multiple case study method is particularly applicable for interpretivist research. In detail, to achieve the goals of this paper, a qualitative approach was adopted through a multiple case study methodology (Eisenhardt, 1989; Gomm et al., 2000), involving firms operating in the retail industry. These firms have adopted big data in business processes. This is deemed an appropriate methodology for this kind of topic because it covers a new domain and because there is limited theory and knowledge about how firms use and manage big data. In this respect, the case study is an appropriate methodology when researchers want to answer questions about the “how” and “why” of a certain topic. Moreover, multiple cases enable the building of a more generalizable and robust theory than single cases (Eisenhardt and Graebner, 2007).

Specifically, the empirical research entailed several semi-structured interviews with marketing managers of each retail firm. The marketing manager was selected as a key respondent because the marketing function is responsible for the data management in retail firms and marketing managers are in a position to understand all the strategic dynamics with the business environment. Interviews were semi-structured to let respondents talk about important aspects not strictly related to those asked (Irvine et al., 2013). This allowed us to get a broader picture of the phenomenon and to discuss some of the themes highlighted with our literature review (Gaur and Kumar, 2017). Data from the interviews were integrated with secondary data from documents, such as business publications and corporate materials, as well as internet-based information. The diversity of data allowed a certain degree of triangulation to improve the reliability of the analysis and strengthened the accuracy of the findings (Jick, 1979). The language of the interviews was Italian; therefore, they have been translated. The nature of this topic required us to gather different types of information about business, strategies and functions, in order to interpret data in a proper way and provide management with useful insights about managerial decisions.

Despite some typical weaknesses of the qualitative approach, case studies have numerous strengths, such as high-conceptual validity, understanding of both the context and process and depth of the analysis (Yin, 2013).

Five organizations operating in the retail industry were selected. To avoid bias, the following steps and procedures were used. First, a convenience selection was made among the best performing organizations in this sector, in order to provide insights deriving from
high-performing benchmarks. They were then contacted to inquire about their availability for participating in interviews and their approach toward big data deployment. Five were available for interviews and were the most interested in this research and topic. Also, these were firms actually deploying big data in processes and activities (at various level). In sum, the following case selection criteria were chosen: the case is represented by a high-performing retail firm, and the case presents actual implementation of big data platforms or initiatives, to some extent. After applying these criteria, one firm was excluded as it had not deployed big data like the other ones.

The four selected firms had deployed big data in their activities at least one year before the interviews. Thus, the firms were surely aware of the mechanisms underpinning big data management and analysis, and their effects were more evident. The choice was made because these firms were active in using big data for several processes and could thus be optimal examples of best practices. The purpose was twofold. First, by using examples coming from best practices, key lessons about management decisions driven by big data could be provided. Second, real examples of big data deployment allowed for highlighting the opportunities and challenges involved with big data deployment, specifically regarding data deployment, data gathering, required competencies and data sharing.

Marketing managers were asked questions regarding general information about business performance, competition, success factors, internal and external opportunities and threats. More specific questions were asked regarding the advantages and challenges of decision-making processes driven by big data.

Data analysis began with considering the mission, vision, values and strategies of the firm, along with the overall history. These data were triangulated with data obtained from the interviews and the results were analyzed. To fill in missing details, when necessary, follow-up correspondence was conducted with the firm via e-mail and/or telephone. To analyze data, an inductive approach was used, following procedures suggested by Miles and Huberman (1994) regarding data reduction, data display and drawing and verifying conclusions. Accordingly, after gathering the data, they were reduced and displayed to identify common paths and develop consistent categories to discuss the findings, drawing and verifying conclusions.

In accordance with the firms, it was decided to provide the analysis as anonymous case studies to prevent any possible misinterpretations due to the open nature of its content (Ben Oumlil, 2013). Firms preferred to remain anonymous to avoid revealing strategic decisions and aspects, especially because of the high concentration of this industry and the limited number of well-performing players in it. Finally, anonymous cases may allow for extrapolating more real information from respondents.

**Results from the case studies**

Overall, it is interesting to highlight that all the four retail firms deeply and steadily use big data analytics to improve processes. All the interviewees stated that big data in the retail industry play a key role, especially in the marketing and logistics functions and that this importance will only increase. The interviews suggest that the use of big data is vital in supporting business strategies for reducing cost structure to win price wars and for differentiating offerings through the segmentation of clients. A further starting point for analysis comes from how data support human decisions. The result of the interviews shows that overall the subjects involved see an ever-increasing role for data in human decisions. If it is true, in fact, that the human element will remain indispensable, it is equally true that in many areas decisions will tend to be based less on experience and intuition, and increasingly on the objectivity of data. Obviously, this does not mean that the human component of the business will disappear, but rather that the tools to support it will be more effective.
**Data deployment**

The key aspect that emerged during the interviews about data deployment pertains to “customer targeting,” that is, the ability to build and understand the customer profile through information gathered from buying approaches. This means being able to understand and analyze the dynamics of clients – their needs, preferences, attitudes – and adapting the offerings accordingly, especially with specific promotions. One response from the interviews was, “Thanks to data, we know almost everything about our regular customers. In this way we can adapt our offering and organization toward meeting their expectations.”

Another application to emerged during the interviews concerns the opportunity to more effectively manage commercial channels, as well as those of marketing, improving the firm’s position within the distribution chain.

Furthermore, the participants spoke of the convenience of exploiting the data collected in order to optimize firm processes (logistic and operational in general) and to reduce operating costs. It is indeed possible to improve the quality level, efficiency and accuracy of deliveries and inventory, but also to implement store efficiency, not only from the point of view of logistics. Human resource management (HRM), for example, can be optimized on the basis of expected and recorded flows within stores, as fixed costs can be monitored constantly and adapted to contingent situations, minimizing waste.

Moreover, it also emerged that the data are often stored and used in order to elaborate detailed plans and forecast budgets related to the opening of new stores, based on models able to anticipate with reasonable accuracy information such as the number of daily entrances expected for economic and financial performance.

**Data gathering**

With reference to the types of data that are collected and used, all firms tend to base their analyses on a mix of real-time surveys and information produced mainly in the previous year. Furthermore, it is interesting to observe that most of the organizations used not only a solid structured database, but also a considerable amount of unstructured data (from e-mails, paper documents, images, comments on social networks, blog posts, etc.). Specifically, among the main sources of value for companies, the former is always used in the analysis of data from customers. Much of this analysis tends to be based on information collected through loyalty cards and extracted from sales data. While the latter is obviously always recorded through normal daily operations, a critical issue may arise for companies that have not yet adopted a data collection strategy through loyalty cards or similar tools.

Furthermore, it is interesting to note that not all data have a practical use at the time they are collected, but they are nevertheless stored with the assumption that they will eventually create value later (so-called “secondary value”). As suggested by one manager: “We collect data for the present but especially for the future. Data accumulation can provide you valuable information useful for future strategies.”

Moreover, the companies involved in this phase are in widespread agreement in considering big data an element of primary importance in order to improve knowledge of and service for customers, so much so that they can soon become the main source to draw on to understand consumers’ needs.

**Required competences**

With reference to the necessary skills for data collection and analysis, the firms involved highlighted the desire to internally process all the information collected, to gain control over data and to store data longitudinally. For this reason, each of them has a special internal division, composed of a team with heterogeneous skills and coordinated in order to extract value from the information in its possession (only one mentioned using both an internal
division and external specialists, while the others claimed to rely solely on an internal system. All four firms, thus, have an internal division able to collect and process data by itself, while affirming that they consider it important for new figures to renew the division in the future. Apart from some issues concerning internal data management, the firms showed great confidence in the ability of their teams to effectively collect, analyze and process the data at their disposal. The trust placed in the internal structures does not mean that they are not attentive to the changes that they will have to undergo in order to prove competitive and keep up with the times. This was highlighted by all the respondents as new professional figures are destined to enter the company context. The predictions range from explicitly citing the data scientist as a new reference profile within the new competitive context to stressing the importance of the heterogeneity of skills that these figures will have to possess, with particular attention to the need for effectively combining information and business skills to provide insights to marketing managers about business decisions. It is no coincidence that respondents also mentioned the increased relevance of the “Marketing Information Manager,” a figure that has existed for years but has emerged vigorously in recent years, combining commercial, IT and marketing skills. Another element that stands out is the increasing attention paid to the application of new technologies to the creation of teams and advanced customer relationship management processes, with the aim of perfecting the experience offered to the consumer and implementing his/her engagement and the related loyalty systems. In addition, many analysts are expected to be able to convert the data into strategic lines and practical applications.

However, although the efforts already made to adapt their internal procedures to the new needs of data analysis are repeatedly stressed, along with the constant attention to the latest trends and to internal and external changes to the company, the respondents revealed the difficulty of discovering and recruiting people characterized by the skills consistent with a big data analyst. Collaborations with the academic world are cited as a way to develop shared projects suitable for increasing knowledge and skills, implementing internal processes and proposing innovative solutions to daily problems, but above all for getting in close contact with the best talents from the university and integrating them into their organization. Not by chance, the only company to demonstrate general confidence in the possibility of finding people with the necessary skills for the business without too much difficulty specifies that this is the result of the close partnerships established with the universities. Moreover, securing a privileged channel with the academic world is proven to be a winning and extremely far-sighted strategy, since it is in the academic world that the skills, now scarce, will be developed, giving shape to the people suitable to successfully manage the “incessant digital flood.”

**Data sharing**

The first noteworthy aspect of data sharing concerns the management of informational capital by companies. Three out of four organizations expressed that they follow a close policy regarding data sharing, suggesting that they prefer not to share data with competitors and partners. This decision seems to be dictated principally by two aspects: a commitment to the protection of consumer privacy (which is given much attention given the risks, reputational in the first place, related to the possible controversies on a legal level) and the desire to retain the enormous value of data. The latter aspect undoubtedly represents one of the greatest strategic dilemmas for companies in this digital era. Companies are, therefore, called upon to carry out a careful analysis of the costs and benefits of sharing data and information, in order to understand whether the benefits deriving from doing so will actually exceed the costs in terms of value dispersion. For the moment, the decision to retain data internally seems to prevail. In this regard, just one organization stated that it would be open to sharing data with competitors and partners to create shared value through relational value chains and in order to strengthen the collaboration between the different parties.
However, as already discussed, examples of sharing information with partners could grow to such an extent that it leads to a shift in the dominant trend among organizations.

Discussion and conclusions

Despite the growing interest in the adoption of digital technologies and big data analytics, a clear picture of the dynamics of the application and benefits of big data deployment is missing in the literature, especially in specific industries, such as retail, one dimension of the service world. In particular, empirical studies on this topic are lacking. Studies in this field mostly consist of recently published conceptual research (Wamba et al., 2015; Wamba and Mishra, 2017). Most of the empirical research is focused on the health industry (Wang, Kung and Byrd, 2018; Wang, Kung, Wang and Cegielski, 2018), with some in the food industry as well (O’Connor and Kelly, 2017).

Consequently, this research aimed to understand how big data deployment transforms organizational practices, thereby generating potential benefits, in the specific service industry of retail.

To fill in the relevant gaps and contribute to the literature, this paper has explored the role of big data in improving the competitive position of retail companies. In particular, an inductive approach by conducting semi-structured interviews has been used. The findings of this research (Table I) highlight the benefits of big data deployment, such as customer targeting (Hofacker et al., 2016; Motamarri et al., 2017) and optimized processes (Lu and Ramamurthy, 2011; Davenport, 2014) as suggested in the literature, but also reduced operating costs; improved quality, efficiency and accuracy of deliveries and inventory; optimized HRM, more detailed plans and more accurate forecast budgets. As an initial implication for theory, the findings clearly indicate the key role of big data for retail firms in achieving competitive advantage. Without big data deployment, retail firms are not able to support business strategies for cost leadership or differentiation (Barney, 1986). As indicated, the four firms included in the analysis use and manage big data in a similar way even though their business strategy is different. In fact, some focus on cost leadership while others emphasize differentiation. Nevertheless, big data management helps firms to achieve and exploit their business strategy. Another theoretical implication regards the link between big data deployment, the competencies required to deploy data analysis and data sharing. Accordingly, it has been demonstrated that big data deployment affects

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Finding</th>
<th>No. firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data deployment</td>
<td>Customer targeting</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>Optimize processes</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>Reduce operating costs</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>Improve quality, efficiency and accuracy of deliveries and stocks</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>Optimize HRM</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>Elaborate detailed plans</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>Forecast budgets</td>
<td>4/4</td>
</tr>
<tr>
<td>Data gathering</td>
<td>Mixed real-time data and stored data</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>Data collected from loyalty cards and sales</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>Use data for a secondary value</td>
<td>4/4</td>
</tr>
<tr>
<td>Competences needed</td>
<td>Heterogeneous skills</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>Searching for new profiles as data scientist, marketing information manager</td>
<td>4/4</td>
</tr>
<tr>
<td></td>
<td>Partnership with universities to acquire and developed the needed profiles</td>
<td>4/4</td>
</tr>
<tr>
<td>Data sharing</td>
<td>Close and tight control of data</td>
<td>3/4</td>
</tr>
<tr>
<td></td>
<td>No data sharing</td>
<td>3/4</td>
</tr>
</tbody>
</table>

Table I. Common findings for all the four cases
different business functions as the need for skilled human resources arises along with the need for data infrastructure, calling for a collaborative approach to leveraging external resources (Ferraris et al., 2017; Scuotto, Santoro, Bresciani and Del Giudice, 2017; Santoro, Bresciani, and Papa, 2018).

This paper, thanks to the case studies, also offers relevant insights and implications for managers. First, it proposes several examples of best practices and opportunities drawn from the case studies that may be useful for managers who are interested in the deployment of big data for business processes. The findings indicate that big data can be used for both internal processes and external opportunities related to customers’ satisfaction and market changes. This is particularly relevant for retail companies, for which increasing competition and the emergence of new business models call for new approaches to data, information and knowledge acquisition in quick and efficient ways. From a business strategy perspective, it is well known that firms always look for ways to reduce costs and increase sales in order to optimize the gross margin contribution. Big data offers the opportunity to work instantly on prices and costs, leveraging the right business strategy and the right time. Accordingly, big data should be thought of not only in terms of analytics but more in terms developing high-level skills that allow for the use of a new generation of IT tools and architectures to collect data from various sources, and to store, organize, extract, analyze and generate valuable insights and share them with key firm stakeholders for competitive advantage.

A key point in this regard is the establishment of partnerships between heterogeneous actors to merge various skills, competences and resources (Tardivo et al., 2017; Vrontis et al., 2017; Del Vecchio et al., 2018). A necessary resource is the human resource, which the interviewed managers suggested was essential. All firms, including those with the best performance, encountered difficulties in hiring talented human resources (Ferraris et al., 2018) who are able to merge data management competences with problem-solving and decision-making skills.

Despite the implications for theory and practice, the paper has several limitations. In particular, a direct relationship between big data deployment and competitive advantage/higher performance cannot be proved. In fact, the paper is based on qualitative evidence of the advantages and opportunities of big data. However, it is believed that the cases presented address key points about strategic exploitative and explorative business processes. Moreover, the firms involved in this research are all high performing, thus suggesting best practices and benchmarks for follower firms. However, future studies could implement quantitative research to assess the impact of big data deployment on a firm’s performance. Another limitation concerns the generalizability of the results. All the firms in this research operate in the retail sector of the service industry. Future studies could focus on the exploration of big data deployment in manufacturing firms.

References


**Corresponding author**
Gabriele Santoro can be contacted at: gabriele.santoro@unito.it

For instructions on how to order reprints of this article, please visit our website: [www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)
Or contact us for further details: permissions@emeraldinsight.com
A bibliometric analysis of research on Big Data analytics for business and management

Lorenzo Ardito
Department of Engineering, Università Campus Bio-Medico di Roma, Rome, Italy

Veronica Scuotto
Research Center, Pole Universitaire Leonard de Vinci, Paris La Défense, France and Department of Management, University of Turin, Turin, Italy

Manlio Del Giudice
Link Campus University Rome, Rome, Italy and Paris School of Business – Rue Nationale, Paris, France, and

Antonio Messeni Petruzzelli
Politecnico di Bari, Bari, Italy

Abstract

Purpose – The purpose of this paper is to scrutinize and classify the literature linking Big Data analytics and management phenomena.

Design/methodology/approach – An objective bibliometric analysis is conducted, supported by subjective assessments based on the studies focused on the intertwining of Big Data analytics and management fields. Specifically, deeper descriptive statistics and document co-citation analysis are provided.

Findings – From the document co-citation analysis and its evaluation, four clusters depicting literature linking Big Data analytics and management phenomena are revealed: theoretical development of Big Data analytics; management transition to Big Data analytics; Big Data analytics and firm resources, capabilities and performance; and Big Data analytics for supply chain management.

Originality/value – To the best of the authors’ knowledge, this is one of the first attempts to comprehend the research streams which, over time, have paved the way to the intersection between Big Data analytics and management fields.

Keywords Bibliometric analysis, Business and management, Business analytics, Big Data analytics, Document co-citation analysis

Paper type Research paper

1. Introduction

With the diffusion of Web 2.0 and Web 3.0 technologies, the advent of the Internet of Things, and the beginning of the Industry 4.0 revolution, an enormous amount of data (i.e. Big Data) about products’ life-cycles, inbound/outbound logistics, product–customer interactions and market needs is available real time (Atzori et al., 2010; Gubbi et al., 2013; Ardito, D’Adda and Messeni Petruzzelli, 2018; Davenport et al., 2012). This has led to the so-called Big Data era. In this era, data are characterized by three characteristics as high volume, variety and velocity, that is, the 3 Vs of Big Data (McAfee and Brynjolfsson, 2012). On the one hand, many opportunities emerge from the possibility to steadily acquire, structure and assess a great amount of data in terms of knowledge generation, decision making and forecasting (Xu et al., 2016; Frisk and Bannister, 2017). On the other hand, the 3 Vs make the process of turning data into valuable knowledge as a very complex task due to storage, standardization, security and data quality issues (Cai and Zhu, 2015; Tankard, 2012). Consequently, many studies have attempted to provide data analysts with novel techniques and more efficient/effective algorithms (e.g. machine learning, recommender systems, etc.), with the aim of[1] solving the above-mentioned shortcomings (Kambatla et al., 2014; Landset et al., 2015; Meng et al., 2014;
Singh and Reddy, 2014). However, these extensive research efforts in improving Big Data analytics techniques and algorithms have not gone hand-in-hand with research discussing the implications of Big Data for business and management. In other words, the extant research on Big Data analytics has especially focused on the technicality underlying related techniques, while remaining disconnected from business and management studies, which have only recently shown a boost in trying to unveil the benefits of Big Data analytics for managerial purposes. This is a relevant gap because despite the exploitation of Big Data, it is one of the main activities that firms can rely on to achieve a sustainable competitive advantage in the next future (Chen et al., 2012), it is useless without elaborating on Big Data analytics with a managerial perspective in mind (Ardito, Messeni Petruzzelli, Panniello and Garavelli, 2018; Sumbal et al., 2017). Indeed, Big Data are useful when they lead to business knowledge that can guide the managerial practice. This issue is exacerbated by the fact that, more in general, the digital economy is challenging traditional economic and business, thus questioning the explanatory power and usefulness of extant management theories (George et al., 2014; Nambisan et al., 2017). In turn, the research on Big Data analytics for business and management is evolving in the absence of a clear theoretical underpinning.

Therefore, considering that research on the technicality underlying Big Data analytics has remained disconnected from the research focusing on the managerial aspects of Big Data and that the (scant) managerial research on Big Data analytics lacks a clear theoretical framework, it is relevant that scholars improve their understanding of how Big Data analytics techniques converge within business processes, and this goal may be better guided if the facets of the literature that lies at the intersection of Big Data analytics and management realms are unveiled. According to the foregoing discussion, the main objective of this research is to scrutinize and classify the literature linking Big Data analytics and management phenomena, so providing a better comprehension of the literature streams that, over time, have paved the way to the intersection between Big Data analytics and management fields.

To do so, an objective bibliometric analysis is conducted (Di Stefano et al., 2010) and supported by subjective assessments based on the previous and current studies on the intertwining of Big Data analytics and management fields. This has allowed us to offer deeper descriptive statistics and document co-citation analysis, being considered necessary to identify, examine and trace the intellectual linkages of a given academic field and guide its future development (Appio et al., 2014). From the document co-citation analysis and its evaluation, four clusters depicting literature linking Big Data analytics and management phenomena emerged: theoretical development of Big Data analytics; management transition to Big Data analytics; Big Data analytics and firm resources, capabilities and performance; and Big Data analytics for supply chain management (SCM). The paper proceeds as follows. Next section introduces the concept of bibliometric analysis. Afterward, the methodology is presented. Then, the results of the bibliometric analysis are showed. Finally, discussion, conclusions and future research directions are provided.

2. Bibliometric analysis
The main objective of this research is to analyze and systematize the various facets of extant literature that lies at the intersection of Big Data analytics and management realms. This objective falls into the broader goal of review studies. So far, two approaches have emerged as the most suitable to pursue these types of goals. On the one hand, there is the subjective approach, which is based on scholars’ interpretation of a given field of research, hence leading to a more qualitative analysis, e.g., a systematic literature review (Tranfield et al., 2003). On the other hand, there is the objective approach, which is instead based on bibliographical and quantitative methods enabled by specific software (Di Stefano et al., 2010; Gaur and Kumar, 2018). Of course, neither the former nor the latter is perfect, and a
combination of the two approaches appears necessary to comprehend the structure and peculiarities of any research domain. Yet, the objective approach is being more and more adopted, especially to examine emerging lines of inquiry that are of interest to different domains (e.g. research on Big Data analytics). Indeed, qualitative approaches are subject to a cognitive bias that is linked to the expertise of the researchers. That is, researchers’ prior experience leads them to pose more emphasis on more familiar domains. Thereby, the probability to undervalue hidden and unexpected connections between diverse literature streams grows, thus limiting the overall value of the literature analysis (Appio et al., 2014).

Accordingly, this study proposes the use of an objective and a subjective approach to examine how the topic of Big Data analytics has been integrated into business and management fields. Specifically, the bibliometric analysis will be adopted as the objective approach. Bibliometrics refers to “the collection, the handling, and the analysis of quantitative bibliographic data, derived from scientific publications” (Verbeek et al., 2002, p. 181). It consists of general descriptive statics (e.g. identifying the main authors, publishing journals, etc.) (Wu and Wu, 2017) and more sophisticated methods like the document co-citation analysis, which is among the most used bibliometric techniques (Fahimnia et al., 2015; Appio et al., 2016; Liu et al., 2015). Its rationale lies behind the fact that any field of study is the result of a cumulative research tradition that can be captured by means of citations; in other words, citations of an article represent the intellectual linkages to certain areas of research. Hence, by examining these intellectual linkages, it is possible to systematically scrutinize relationships among documents contributing to the development of a defined research field (Di Stefano et al., 2010). The output of the document co-citation analysis is often a map that relies on network theory to explore the data structure (Liu et al., 2015). Such map helps scholars to identify relevant research domains and assess the extent to which they are connected. Notably, a document co-citation map consists of a set of nodes, representing cited documents and a set of links, representing the co-occurrence of nodes in the reference list of papers upon which the map is based. Documents are co-cited if they appear together in the reference lists of a publication. For instance, Documents A and B are co-cited if both are cited by a Paper C. The more publications co-cite A and B, the stronger the relationship between A and B. This implies that they belong to a similar subject area (Hjørland, 2013; Fahimnia et al., 2015). A qualitative assessment will complement such analysis. Eventually, by relying on document co-citation maps and a qualitative assessment of the maps, the literature linking Big Data analytics and management phenomena will be scrutinized and classified.

3. Methodology
To design a literature review (regardless its qualitative or quantitative nature), build the bibliography and derive reliable results, an “iterative cycle of defining appropriate search keywords, searching the literature, and completing the analysis” is required (Fahimnia et al., 2015, p. 103; see also Tranfield et al., 2003). Specifically, to conduct a bibliographic (co-citation) analysis, Fahimnia et al. (2015) suggested a five-step methodology for data collection and comprehensive evaluation of the selected field of study. It entails the definition of the database to query and the search strategy (i.e. the search terms to adopt), the screening of the initial search results, the refinement of the search results, the development of initial descriptive statistics and detailed bibliometric data analysis (i.e. the co-citation analysis). In the following, we will describe Steps 1–3, while Steps 4 and 5 (revealing the results) are presented in Section 4.

Since our aim is to scrutinize the literature on Big Data analytics for business and management, we sought to collect the list of published articles related to the adoption/implications of using Big Data analytics techniques for business and managerial purposes. In detail, the identification of the theoretical underpinnings of these articles is in the focus of this study. Thereby, an analysis restricted to top-tier journals may help to understand
how a research domain is actually evolving from a theoretical perspective. Accordingly, journals provided with the impact factor, which is a well-established indicator of a journal’s high quality (e.g. Meier, 2011), are considered. Moreover, we did not include technical journals that are more focused on the development/performance of specific Big Data analytics algorithms, which is out of the scope of this research. For these reasons, we relied on the Thomson ISI Web of Science (WoS) database as the data source (e.g. Di Stefano et al., 2010; Tan et al., 2014) since, unlike Scopus, DBLP, WorldCat, IEEE Xplore, etc., it allows directly searching for keywords in journals pertaining to the Social Science Citation Index (SSCI). The SSCI, in fact, includes journals with impact factor and referring only to the domains of social science as business and management. Nonetheless, we did not further restrict the list of journals for investigation (e.g. through cut-off levels of the impact factor or publishing period) given the novel nature of the topic. Still, conference proceedings, book chapters and all publication types other than peer-reviewed articles were left out to assure an examination of high-quality publications (e.g. Liu et al., 2015; Meier, 2011).

To identify search terms and, in turn, retrieve articles for inclusion, we employed a screening routine (Appio et al., 2014; David and Han, 2003) that allowed us to identify the best search strategy. We started by querying the WoS database with the search term “Big Data analytics” in the field “Topic,” thus gathering 223 articles. Afterward, we revised the query string in such a way that the search term “Big Data analytics” must be associated with, at least, one of the following terms: “firm,” “performance,” “innovation,” “strateg*,” “capabilit*,” “resource*,” “knowledge,” “absorptive,” “management,” “organization*” or “business.” Indeed, such terms are strongly related to business and management fields. This procedure yielded a sample of 178 articles. We scrutinized the 45 articles not included in the second search strategy, and they actually departed from the topic under investigation since no direct implications for business and management fields were provided. As a further check, we replicated the articles search by dividing the terms “Big Data” and “analytics.” That is, we queried the WoS database considering, first, the search string (“Big Data” AND “analytics”) and, then, the search string (“Big Data” AND “analytics”) AND (“firm” OR “performance” OR “innovation” OR “strateg*” OR “capabilit*” OR “resource*” OR “knowledge” OR “absorptive” OR “management” OR “organization*” OR “business”).

These two search strategies led to 613 and 478 articles, respectively. As in the previous case, the inclusion of the terms “firm,” “performance,” etc. is more conservative and consistent with the aim of the paper. Regarding the separation between the terms “Big Data” and “analytics,” we found that this is a better approach since too many relevant articles were excluded when the unique term “Big Data analytics” is adopted. Eventually (“Big Data” AND “analytics”) AND (“firm” OR “performance” OR “innovation” OR “strateg*” OR “capabilit*” OR “resource*” OR “knowledge” OR “absorptive” OR “management” OR “organization*” OR “business”) was considered the final search string to adopt, and the number of resulting articles considered for the bibliometric analysis was 478.

4. Results of the bibliometric analysis
4.1 General descriptive statistics
Based on general information of the 478 retrieved articles, as a part of the bibliometric analysis, initial descriptive statistics will be provided. Figure 1 reveals the novel nature of the topic, in that the oldest articles are published in 2012 and the number of publications is proven to double every year. The number of articles for 2018 is incomplete due to the fact that the data collection ended at the beginning of June. However, it is interesting to note that it is already close to the number of articles published in 2016.

Furthermore, it is worth mentioning that several journals have contributed to the publication of the 478 articles, which reveals how this topic is widespread in the extant social science literature. Table I reports the journals that have published more than five
articles in total across years, with *Journal of Business Research*, *Sustainability*, *MIS Quarterly*, *Business Process Management Journal*, *Decision Support Systems*, *International Journal of Information Management* and *Technological Forecasting and Social Change* as the most favorable outlets for the topic under investigation. Only recently, other journals (not reported in the table) have started to welcome articles that connect Big Data analytics and management fields (e.g. *Management Decision*, *Human Resource Management Journal*, *IEEE Transaction on Engineering Management* and *Educational Technology & Society*).

Overall, the journals cover several research areas, still with implications for business and management. This may reveal how the theme of Big Data analytics interests several research streams and has probably promoted the cross-fertilization of diverse fields. To dig into this issue, Figure 2 depicts the research areas with more than ten published articles. Many papers appear in more than one category, which explains why the sum of articles considering all research areas is greater than 478. Among the most relevant areas there is, of course, *Business Economics*, but also *Computer Science*, *Operations Research and Management Science* and *Engineering*, thus further suggesting the pervasiveness of the topic under investigation in social science.

Table II outlines top contributing authors (i.e. those with no less than four articles) and the number of articles they authored or co-authored. Yet, there is a long tail of non-reported authors with only one article, thus reflecting that very few scholars have specialized at studying Big Data analytics with a managerial lens. Considering, instead, the authors reported in Table II, some of them co-authored same articles (e.g. Dubey *et al.*, 2018), and their research mainly focuses on Big Data analytics in the context of SCM and business strategy.

Finally, Table III and Figure 3 delve into the affiliations of authors that have authored the 478 sample articles. Table III presents the top performing organizations in terms of number of papers contributed and their geographic location. Comparing this list with the top contributing authors in Table II, we observed that top performing organizations are represented by the more prolific authors. From a geographical perspective, most of the top performing organizations are located in the USA, which hints that US-based organizations lead the research activities that lie at the intersection of Big Data analytics and management domains. This is confirmed by Figure 3 that reports countries with more than ten publications. Indeed, the USA is associated with more than 100 publications (217 articles), followed by England (77 articles) and, interestingly, the People’s Republic of China (62 articles). Except for England, the other European countries appear to be not so involved in the considered line of inquiry.
4.2 Document co-citation analysis

4.2.1 Overview of the method. As discussed in Section 2, documents that are cited collectively represent the intellectual base of a knowledge domain. This characteristic allows document co-citation analysis to inform about the cumulative movement and micro-structure characterizing a certain discipline or research domain as well as the related hidden micro-level connections between diverse knowledge dots (Wagner, 2008).
which may ultimately unveil co-citation clusters representing the diverse streams of the literature, the discipline or research domain reflect (Hjørland, 2013; Bornmann and Leydesdorff, 2015). Accordingly, this research carried out a document co-citation analysis applied on the references of the sample articles, with the ultimate aim of identifying main clusters capturing the bibliometric landscapes of the research about Big Data analytics for managerial purposes. Afterward, by relying on a subjective approach to the objective selection of clusters, this first step was complemented with a screening of the recognized clusters through their aggregation around some key research streams (Chen et al., 2010).

Indeed, the objective approach is no substitute for careful reading and fine-grained content analysis of the results obtained (White and McCain, 1998).
The software VOSviewer version 1.6.8 was employed. VOSviewer generates maps using the VOS mapping and VOS clustering techniques. These are novel, alternative techniques to the Multidimensional Scaling (MDS) approach (van Eck and Waltman, 2010; Waltman et al., 2010). VOS techniques mirror the MDS approach in terms of aim – i.e., “locating items in a low-dimensional space in such a way that the distance between any two items reflects the similarity or relatedness of the items as accurately as possible” (Appio et al., 2014, p. 628). However, differently from MDS, which is based on the calculation of similarity measures like the cosine and the Jaccard indexes, VOS adopts a more appropriate procedure for normalizing co-occurrence frequencies (e.g. van Eck and Waltman, 2007; van Eck et al., 2006), namely, the association strength (AS$_{ij}$)[2]. Specifically, the AS is calculated as follows:

$$AS_{ij} = \frac{c_{ij}}{w_i w_j},$$

which is “proportional to the ratio between on the one hand the observed number of co-occurrences of $i$ and $j$ and on the other hand the expected number of co-occurrences of $i$ and $j$ under the assumption that co-occurrences of $i$ and $j$ are statistically independent” (van Eck and Waltman, 2010, p. 531). On the basis of this index, VOSviewer locates items (documents in our case) in a map after minimizing the weighted sum of the squared distances between all pairs of items. The LinLog/modularity normalization was performed (Noak, 2007; Appio et al., 2016).

4.2.2 Results of the document co-citation analysis. Figure 4 shows the salient map structure of co-cited references derived from the citing behavior of authors who have published the sample articles. In the map, we highlighted the most relevant co-cited pairs, i.e., documents co-cited more than five times (Appio et al., 2014). The bigger the node, the more the citations received by a document; the thicker the link, the more the connected nodes have been co-cited. The presented map shows 216 nodes, 500 links and 4 clusters. Actually, the initial number of clusters was 6, but we conducted a qualitative assessment of clusters on the basis of a content analysis, which led to the 4 identified clusters. Due to the high number of papers in each cluster, we carefully present the contents of the lead papers. The lead papers of each cluster are shown in Table IV, while Table V reports the total number of papers in each cluster and respective author-provided label as a result of the content analysis.

The first cluster, “Theoretical development of Big Data analytics,” refers to those papers aiming at clarifying what “Big Data” and “Big Data analytics” mean. They depart from specific managerial aspects and attempt to identify, classify and systematize the literature on Big Data and Big Data analytics. These studies include a multidisciplinary analysis of
Figure 4. Co-citation map of sample articles' references.
the phenomenon of Big Data. For instance, Mayer-Schönberger (2015) developed a research on the use of Big Data in the medical sector. boyd and Crawford (2012) were more focused on the impact of the Big Data on the technological, social and academic scenario. They questioned how Big Data can be a useful tool to generate more knowledge. On this basis, Provost and Fawcett (2013) paid more attention to the intertwining of the different disciplines related to the Big Data and the data-driven decision making. In turn, they explored the concept of the data science, which is “is a set of fundamental principles that support and guide the principled extraction of information and knowledge from data” (p. 52). Such extraction can create a great value for businesses. However, the value of Big Data cannot be measured if advanced technologies are not used. Chen et al. (2014) conducted a study on the four stages of the value chain of Big Data, i.e., data generation, data acquisition, data storage and data analysis by employing enterprise management tools smart grid and social media networks, among others. Although the benefits provided by Big Data, firms have also to overcome different challenges such as data visualization, data capture, data analysis and data storage (Chen and Zhang, 2014). Additionally, Lazer et al. (2014) argued the importance of using the terminology of “all data revolution” rather than “Big Data revolution” because of the inclusion of small data as well. Indeed, they assumed that small data can be more effective than Big Data, involving a non-traditional data collection. On a different lens, Hashem et al. (2015) focused on cloud computing to better perform a data analysis, where “cloud computing is one of the most significant shifts in modern ICT and service for enterprise applications and has become a powerful architecture to perform large-scale and complex computing. The advantages of cloud computing include virtualized resources, parallel processing, security, and data service integration with scalable data storage” (p. 99). Finally, Gandomi and Haider (2015) attempted a definition of Big Data by assembling insights from academics and practitioners. In turn, they described Big Data as a form of structured, semi-structured and unstructured data, distinguished also in size, speed and variety from which they grab insights for businesses.

The second cluster, “Management transition to Big Data analytics,” refers to explorative studies theorizing about the need to create a link between the Big Data analytics and management fields. Indeed, Chen et al. (2012) showed the relevance of the business

<table>
<thead>
<tr>
<th>Cluster 1 (red bubbles)</th>
<th>Cluster 2 (blue bubbles)</th>
<th>Cluster 3 (green bubbles)</th>
<th>Cluster 4 (yellow bubbles)</th>
</tr>
</thead>
</table>

Table IV.
The lead papers of each cluster

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number of documents</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>52</td>
<td>Theoretical development of Big Data analytics</td>
</tr>
<tr>
<td>2</td>
<td>78</td>
<td>Management transition to Big Data analytics</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td>Big Data analytics and firm resources, capabilities and performance</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>Big Data analytics for supply chain management</td>
</tr>
</tbody>
</table>

Table V.
The four major clusters
intelligence and analytics (BI&A) for businesses. Especially, they explored three different versions of BI&A, 1.0, 2.0 and 3.0, in order to discuss the data-related issues occurred within the organizational environment. On this lens, McAfee and Brynjolfsson (2012) stressed the significance of Big Data in the decision-making process. Manyika et al. (2011) explored more the phenomenon of the competition derived from the Big Data, already argued by Davenport (2006), who also studied the dark side of Big Data analytics in managing consumers’ data which give a competitive edge “that your competitor won’t be able to match” (Davenport and Harris, 2007, p. 3). Hence, firms, mainly corporates, are more prone to develop a data-driven strategy (Barton and Court, 2012), which calls for a new player in firms, as represented by the data scientist, who is an expert in Big Data and seeks to discover new things in the digital world (Davenport and Patil, 2012).

The third cluster, “Big Data analytics and firm resources, capabilities and performance,” reflects the research that has first attempted to provide empirical insights regarding the implications of the adoption of Big Data analytics for managerial purposes. This, hence, discusses the Big Data analytics for achieving better firm performance or how they help firms to acquire more resources and/or develop novel capabilities. Accordingly, this cluster is based on the very well-established studies by Barney (1991) and Teece et al. (1997) about “firm resources and sustained competitive advantage” and “dynamic capabilities and strategic management,” respectively. In turn, these studies spurred further academic conversations in the field of Big Data analytics. For example, Bharadwaj (2000) examined the link between IT capability and firm performance. Yet, the most relevant studies on the topic of Big Data and resource capability were offered starting from 2011. For instance, LaValle et al. (2011) questioned how a firm can achieve the best value, especially with respect to traditional analytics (Davenport et al., 2012). Given the fact that companies are firmly aware of the value of Big Data analytics, they tend to prioritize the enhancement of knowledge within their organizations by using three different levels of capabilities: aspirational, experienced and transformed. Instead, Murdoch and Detsky (2013) explored the adoption of Big Data to the health care system to better support practitioners’ delivery service. Alongside, Sharma et al. (2014) provided a deeper analysis of the process of allocating and managing resources through the use of business analytics which, in turn, can enhance business performance. Concluding, Erevelles et al. (2016) emphasized the powerful benefit in getting insights of consumers from Big Data and, thus, in better implementing marketing strategies. All in all, this cluster complements recent studies on governance structure, resource management and internalization processes that give a wide overview on the “firm resources, capabilities and performance” (e.g. Gaur and Kumar, 2018; Singh and Delios, 2017; Contractor et al., 2016; Singh and Gaur, 2013).

Finally, while Cluster 3 covers a wide range of managerial issues in diverse sectors (e.g. health), Cluster 4, “Big Data analytics for supply chain management,” hints that, in the wide realm of management studies, the research area related to operations and SCM is the most interested in analyzing the implications of the adoption of Big Data analytics. Accordingly, Waller and Fawcett’s (2013a) study was among the first in highlighting the revolution of the SCM caused by Big Data, predictive analytics and data science. Additionally, by employing a $2 \times 2$ matrix of prediction and explanation, they argued the significance of predictive analytics. In this line, Fosso Wamba et al. (2015) offered a longitudinal case study showing how to enhance operations delivery by employing IT capabilities and multiple sources of data, while Schoenherr and Speier-Pero (2015) offered an evaluation of the current state of SCM predictive analytics via a large-scale survey. Their study showed an involvement of the university in the SCM predictive analytics in forming the new experts known as data scientist. Furthermore, still following the research mainstream on the SCM, Hazen et al. (2014) proposed a method to monitor and control data quality by exploring different theories such as
knowledge-based view, organizational information processing view and system theory, among others. Relatedly, Tan et al. (2015) examined the supply chain innovation capabilities by analytic infrastructure relied on the deduction graph technique, and Wang et al. (2016) conducted a research on Big Data logistics and SCM, stressing the relevance of understanding Big Data for managers. Finally, a more generic study is provided by Dutta and Bose (2015), who analyzed the Big Data project developed in a manufacturing company, Ramco Cements Limited in India, demonstrating that innovative visualization methods, a pervasive organizational direction toward data decision-making process and a cross-function project team are the key aspects for the success of a Big Data project.

5. Final remarks and directions for future research

The paper offers a bibliometric analysis, supported by a qualitative approach, of the concept of Big Data analytics showing the different nuances of this concept mainly referring to the managerial literature.

Big Data analytics has the potential to revolutionize the art of management. Despite the high operational and strategic impacts, there is a paucity of empirical research to assess and prove the business value of Big Data. Additionally, previous research was more focused on big corporates rather than small to medium enterprises (SMEs). This has resulted in a need for more analysis on different firms’ size by both a qualitative and quantitative approach. Stated differently, extant research should be enlarged by taking into consideration SMEs. Likewise, since data can be more easily gathered worldwide, internationalization issues in the context of Big Data analytics have yet to be examined. The novel side of these topics is related to the dynamic nature of digital transformation, which is involving all kind of companies – from micro to big firms. Moreover, there is a new big dilemma: how SMEs are embracing and managing such big transformation even though their lack of resources or also on a different note, how the big corporates have embedded this Big Data revolution in their process of internationalization. Additionally, looking into the aforementioned studies focused on SCM, the concept of networking emerged. It can spur further research in exploring the relationship between Big Data, academic ecosystem and SCM. Furthermore, it is also connected to another theme that is firms’ resources, capabilities and performance. This topic appears to be more developed in the last decade. In fact, papers are mainly dated from 2006. Notwithstanding, there are a few research works dated in 1991 (Barney, 1991) and 1997 (Teece et al., 1997) that are selected for their relevant studies on firms’ resources, capabilities and performance, which were inspiring further research on this topic and Big Data. The majority of the papers were coming from the USA, although there is an increasing number of papers analyzed from non-western country authors (Chen et al., 2014; Chen and Zhang, 2014; Wang et al., 2016).

Despite each cluster has its clear own specificities, it is interesting to note that clusters are linked among them by certain papers that we can define as bridges. In particular, there are few papers that are especially relevant within each cluster – being at the center of very dense co-citation networks – and between clusters – acting as connectors among them. Figure 5 reveals that the paper by Chen et al. (2012) is the most representative in this sense, but also the papers by Fosso Wamba et al. (2015), LaValle et al. (2011) and Mayer-Schönberger (2015) connect all four clusters while playing a crucial role in their own.

A great interest in Big Data derived from the works related to the SCM, where Big Data is more and more affecting firms’ performance. This stimulates further research employing a quantitative approach and also extending this analysis to the non-western countries, which have barely been examined (Wang et al., 2016).

In a nutshell, this paper presents an interpretive framework that analyzes the definitional perspectives and the applications of Big Data analytics in the management field. The paper also provides a general taxonomy that helps broaden the understanding of Big Data and its role in
fostering business value. The synthesis of the diverse concepts within the literature on Big Data provides deeper insights into achieving value through Big Data strategy and implementation. As with most studies, this research offers interesting insights, but it is also affected by some limitations. For instance, a multidisciplinary approach would request for a different bunch of keywords such as “multidisciplinary,” “multiculturalism” and “computing.” Therefore, future research can be deeper and examine this scenario and enlarge the pool of papers. Alongside, there are other chances for a content analysis which may show other key aspects of the current research topic on Big Data analytics.

Notes
1. Polytechnic University of Bari, Viale Japigia 182 – 70126 Bari.
2. Please refer to Van Eck and Waltman (2010) for an extensive discussion of the advantages of the association strength over other similarity measures, such as the cosine and the Jaccard indexes.

References


**Corresponding author**
Veronica Scuotto can be contacted at: veronica.scuotto@gmail.com

For instructions on how to order reprints of this article, please visit our website: 
[www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: permissions@emeraldinsight.com
Talent management under a big data induced revolution

The double-edged sword effects of challenge stressors on creativity

Kaidi Zhang
Department of Innovation, Entrepreneurship and Strategy,
School of Economics and Management, Tsinghua University, Beijing, China

Xiao Jia
Department of Innovation, Entrepreneurship and Strategy,
School of Economics and Management, Tsinghua University, Beijing, China and
Research Center for Technological Innovation, Tsinghua University,
Beijing, China, and

Jin Chen
Research Center for Technological Innovation, Tsinghua University,
Beijing, China

Abstract
Purpose – The emerging natures of big data – volume, velocity, variety, value and veracity – exert higher stress on employees and demand greater creativity from them, causing extreme difficulties in the talent management of organizations in the big data era. The purpose of this paper is to explore the effect of challenge stressors on creativity and the boundary conditions of the relationship.

Design/methodology/approach – Multisource data were collected including 593 followers and their 98 supervisors from organizations that are confronting a big data induced management revolution. Hierarchical regression analysis and bootstrapping analysis were used to test the mediation and moderation mechanism.

Findings – The results showed that job burnout mediated the negative relationship between challenge stressors and creativity and that this indirect effect was attenuated by an employee’s core self-evaluation (CSE) and servant leadership. In contrast, whether work engagement mediated the relationship between challenge stressors and creativity was contingent on the level of an employee’s CSE and servant leadership. Specifically, the mediating effect was significant only when an employee’s CSE or servant leadership was high.

Originality/value – The results contribute to our understanding of the relationship between challenge stressors and creativity in the big data era. Specifically, relying on the job demands–resources model, this study empirically opens the “black box” between challenge stressors and creativity by exploring two opposing intermediate mechanisms. In addition, this study reveals boundary conditions by investigating dispositional and contextual factors that can accentuate the positive effect while attenuating the negative effect of challenge stressors on employee creativity.

Keywords Creativity, Challenge stressors, Big data era, Management revolution, Talent management transformation

Paper type Research paper

Introduction
We are drowning in data (Gobble, 2013). In changing how we see the world, big data not only is a new tool for technology innovation (Strawn, 2012) but also claims to introduce a new management revolution era (McAfee and Brynjolfsson, 2012; Wamba et al., 2015). Big data adaptation can come from internal motives, which include management transformation (Wamba et al., 2015), or external motives, such as digitalized competition and technology challenges (El-Kassar and Singh, 2018). Needs and opportunities coexist when institutions and policy changes interact in nonlinear and unpredictable ways (Singh and Gaur, 2018). Currently, organizations can achieve the competitive benefits of...
transitioning to big data, enabling organizations only when they can manage the challenge stressors from this revolution effectively and creatively (McAfee and Brynjolfsson, 2012). Some scholars (El-Kassar and Singh, 2018; Wamba et al., 2015, 2018; Acharya et al., 2018) describe big data’s key differences in terms of 5Vs: “Volume”: the large amount of data that companies can utilize from numerous and independent sources. “Velocity”: the speed of data creation makes it possible for a company to analyze real-time or near real-time information. “Variety”: data are generated from various sources and formats, such as images posted on social networks, cellphone signals and readings from sensors. “Value”: the importance of extracting economic benefits from the available big data. “Veracity”: the significance of data quality and data sources trust level.

Indeed, an increasing number of stressors and the increasing requirements for creativity are two key trends in most organizations in the big data era. First, a big data context undoubtedly exerts more stressors – which indicate an individual’s perceptions of workload, time pressure and job complexity (LePine et al., 2005) – on employees, given data’s enormous quantity, rapid change, great variety, high uncertainty and large ambiguity. Meanwhile, the big data era requires a higher level of creativity from employees (Gobble, 2013). Specifically, employees must be creative to manage large quantities of data, know how to collect, store, organize, analyze and share big data and generate novel ways to exploit the value of a big data set. In all, in the context of the big data era induced management revolution, employees are faced with an increasing number of challenge stressors and are required to exhibit higher levels of creativity, which is the key for organizations to achieve a competitive advantage in the big data era. From a talent management transformation perspective, this study explores whether challenge stressors have a positive or negative effect on employee creativity.

As data become cheaper, talent management is crucial (McAfee and Brynjolfsson, 2012). The majority of previous studies focused on value creation from big data, such as creating transparency, enabling experimentation to discover needs, segmenting population to customize action, supporting human decision making with automated algorithms, innovating new business and knowledge co-creation (Acharya et al., 2018; El-Kassar and Singh, 2018; Wamba et al., 2015). However, few of them focus on another important aspect: the management of talents in the big data era. In addition, the relationship between challenge stressors and creativity also remains unclear. Generally, it is believed that challenge stressors have a positive effect on creativity (Lepine et al., 2005). Individuals under challenge stressors are highly motivated to be creative, as they perceive a positive assumption that the more efforts required to deal with challenge stressors, the more mastery and personal growth is expected, and, subsequently, valuable achievements can also be expected (Lepine et al., 2005; Cavanaugh et al., 2000; Crawford et al., 2010). However, the opposite view also exists. Some researchers argue that challenge stressors are believed to impair creativity, as they “tax too highly the limited cognitive resources necessary for creativity” (Sacramento et al., 2013, p. 142). In addition, a curvilinear relationship (Gardner, 1986) and a non-significant relationship (De Dreu and West, 2001) are also identified.

The current study attempts to extend this inquiry by focusing on the potential competing intervening mechanisms underlying challenge stressors: creativity relationships in the big data era. Specifically, we argue that the mixed effects can be accounted for by the fact that challenge stressors might operate through two parallel but competing pathways to influence individual creativity. Thus, a positive intermediate mechanism would operate side by side with a negative mechanism to translate challenge stressors into creativity. The job demands–resources model (JD–R model) (Bakker and Demerouti, 2007; Bakker et al., 2014; Tetrick and Winslow, 2015) offers a theoretical framework for an understanding of the complex intermediate dynamics. Drawing on the JD–R model, the opposing effects of challenge stressors on employee creativity should be mediated by job burnout (resource-depleting mechanism) and work engagement (motivational mechanism), respectively.
After establishing the competing mechanisms, other questions include how to amplify the salutary effect using the motivational pathway and how to buffer the detrimental effect using the resource-depleting pathway. Following the JD–R model, we argue that factors that can provide or replete one with sufficient resources accentuate the positive pathway and attenuate the negative pathway, in that resources can help one either reduce job demands and the associated physiological and psychological costs, or stimulate personal growth, learning and development (Bakker and Demerouti, 2007). In the current study, we focus on both personal resource and job resource (Tetrick and Winslow, 2015), which can be operationalized by core self-evaluation (CSE) (Judge et al., 2003) and servant leadership (Greenleaf, 1970, 1977), respectively. Examining the moderating effects of two types of resources can not only provide an understanding of the interaction dynamics of challenge stressors and resources on employee burnout, motivation and the more distal behavioral outcome, creativity, but more importantly provide invaluable insights for practitioners to manage and intervene with increasing challenge stressors.

As depicted in Figure 1, this study aims to address two basic questions on talent management transformation in the big data era: How do employees’ challenge stressors exert double-edged sword effects on creativity? What factors can augment the positive effect but attenuate the negative effect? By examining these two questions, the current study contributes to the literature in three ways. First, we highlight an important talent management issue in the big data era, focusing on how employees’ challenge stressors influence their creativity. Second, by investigating the competing intermediate mechanisms, we reconcile the contradictory insights from the previous studies. Third, based on the JD–R model, we explore the moderating roles of individual CSE and servant leadership, which can buffer the negative effect but amplify the positive effect of challenge stressors on creativity. By doing so, we extend the understanding of both the literature and practice of intervening and managing challenge stressors. In essence, we respond to recent calls for adopting an interactive perspective to investigate how creativity is formed (Zhou and Hoever, 2014).

**Literature review and hypotheses**

**Big data induced management revolution, challenge stressors and creativity**

Big data is not only a management practice revolution (McAfee and Brynjolfsson, 2012) but also a management research revolution with increased attention from academic literature (Wamba et al., 2018). Previous studies have carried out a great deal of research on the value creation from big data. For example, El-Kassar and Singh (2018) developed a holistic model and found that big data and predictive analytics positively influence organizational performance. Dubey et al. (2018) found that big data analytics capability has a positive significant effect on supply chain agility and competitive advantage based on data from 173
organizations of automotive components manufactures in India. Wamba et al. (2018) found that the overall information quality in big data analytics also has a significant positive impact on firm performance using data from 302 business analysts in France and the USA. Acharya et al. (2018) found that big data can assist in knowledge co-creation based on data from four fashion retailing organizations. Though previous studies find that big data could create transparency, enable experimentation to discover needs, segment population to customize action, support human decision making, innovate new business and co-create knowledge (Acharya et al., 2018; El-Kassar and Singh, 2018; Wamba et al., 2015), few of them focus on the management of talents in the big data era proposed by McAfee and Brynjolfsson (2012). Talent management is crucial in an organization, as data become cheaper and more accessible and must differ from their traditional aspects, as seen in the remarkable “5 Vs” that characterize the big data era: volume, velocity, variety, value and veracity, which change everything (McAfee and Brynjolfsson, 2012; Acharya et al., 2018).

The big data context makes individuals experience more workplace stressors due to the enormous quantity, rapid change, great variety, high uncertainty and large ambiguity of data. Recent scholars have applied the challenge–hindrance stressor framework to account for the effects of stressors, as they argue that the nature of stressors is critical to determining their influence (LePine et al., 2005). Specifically, hindrance stressors refer to individuals’ perceptions of their work environment in terms of the level of demands, such as role conflict, role ambiguity, politics, red tape and job insecurity, while challenge stressors indicate individuals’ perceptions of their work environment in terms of the level of demands such as workload, time pressure, job complexity and responsibility (LePine et al., 2005). In a management revolution, substantial changes and challenges are prevalent, and these can be regarded as challenge stressors (Bose, 2013; Tetrick and Winslow, 2015). The challenge stressors from the big data induced revolution contain the following aspects. Decision making: decision maker should embrace evidence-based decision making, from data analysis, exploitation and translation to valuable information for business decisions (El-Kassar and Singh, 2018; Pauleen and Wang, 2017; Wamba et al., 2015). Human resources management: data scientists and computer scientists are the sexist job currently, as the ability to analyze and visualize unstructured and large quantities of data is becoming crucial and rare (Davenport, 2012; Gobble, 2013; El-Kassar and Singh, 2018; McAfee and Brynjolfsson, 2012; Wamba et al., 2015). Technology: individuals should be able to handle the hardware, the open-source software, as well as new requirement for the volume, variety and velocity of big data sets (Boyd and Crawford, 2012; Davenport, 2012). In addition, work habit and organization change: individuals should be more open-minded rather than act alone, treasure what they have more than how they think, treasure correlation more than causation, and perform customized actions (Barton and Court, 2012; Wamba et al., 2015); individuals should also be able to recognize redundant and inaccurate data and enhance the ability of data governance (Schroeck et al., 2012). The big data era has triggered a management revolution, which highlights a concern for challenge stressor management for employees.

On the other hand, in a management revolution (McAfee and Brynjolfsson, 2012), the big data era requires creativity (Gobble, 2013) if individuals want to survive and reap the benefits of the big data era (Anderson et al., 2014; Pauleen and Wang, 2017; Wamba et al., 2015). First, compared with traditional jobs, the big data era introduces new work content. Individuals should be creative in managing large quantities of data, know how to collect, store, organize and analyze the data and their related new accesses and acquisitions (El-Kassar and Singh, 2018). Second, work behaviors are also expected to change. Individuals must begin innovating new ways to exploit the value of big data sets. Individuals should be creative in data-intensive activities and leverage big data to streamline processes and create efficiencies which may innovate general ways of working (Gobble, 2013). In addition, talent competition is
essentially creativity competition. As data scientists are the major players in the big data era, they are more motivated to create new products and services, rather than only support internal decision making (Davenport, 2012).

In the context of the big data era induced management revolution, the intensive challenge stressors and increasing creativity requirements trigger crucial questions about talent management. In the current study, we investigate how challenge stressors and creativity are related and propose two competing intervening mechanisms to explain this relationship based on the JD–R model.

**JD–R model**

One core assumption of the JD–R model is that all job characteristics can be classified into two general categories: job demands and job resources (Bakker and Demerouti, 2007; Bakker et al., 2014; Demerouti et al., 2001). Job demands denote those physical, social or organizational aspects of jobs that require sustained physical or mental efforts which in turn are associated with certain physiological and psychological costs, while job resources refer to those aspects of jobs that are functional: achieving work goals, reducing job demands and their associated physiological and psychological costs or stimulating personal growth and development (Demerouti et al., 2001, p. 501).

Drawing on the JD–R model, we find that job demands and resources can exert an influence on an individual’s well-being and job outcomes through two psychological processes: a resource-depleting process (Bakker et al., 2003) and a motivational process (Bakker and Demerouti, 2007; Bakker et al., 2014). Specifically, job demands are theorized to deplete an individual’s energy and resources, trigger burnout and job burnout, undermine motivation and work engagement and dampen one’s health, well-being and performance-related outcomes. In contrast, job resources help to replete individuals with resources, and fuel their high work engagement, motivation and enjoyment, which, in turn, is positively associated with benign individual and organizational outcomes (Bakker et al., 2003, 2014; Nahrgang et al., 2011). In addition to triggering the dual processes independently, job demands and resources can have joint and interactive effects on individual well-being, and motivational and job outcomes (Bakker et al., 2014). According to the model, job resources would buffer the negative effects of job demands on burnout and strain because employees with a large amount of resources are able to cope better with job demands. Meanwhile, job demands might enhance the positive effects of resources on motivation because job resources become more salient and valuable when employees are confronted with a high level of job demands (Bakker and Demerouti, 2007; Bakker et al., 2014).

In this study, all our hypotheses are based firmly on the JD–R model. We first provide an account of the inconclusive relationship between challenge stressors and creativity by arguing that job burnout (i.e. resource-depleting mechanism) and work engagement (i.e. motivational mechanism) work as parallel but competing intervening variables in the linkage. Further, we draw on the JD–R interaction assumption to theorize about the moderating roles of CSE (internal resources) and servant leadership (external resources) in amplifying the motivational process but buffering the resource-depleting process.

**Job burnout as a mediator**

Higher levels of challenge stressors caused by the big data context lead to job burnout. According to the JD–R model, when individuals appraise their work requirements as resource-depleting, negative reactions occur, such as sleep problems, fatigue and strain. As challenge stressors include higher workload, tougher job demands, and a higher level of job complexity, individuals expend substantial psychological resources to deal with them (Cavanaugh et al., 2000; Cox, 1985), further inducing job burnout. Furthermore, coping with the big data induced management revolution taxes resources (LePine et al., 2005).
As overloaded job demands in the management revolution tax an individual’s extra emotional and cognitive resources, individuals are expected to perceive resource depletion, which also results in job burnout (Crawford et al., 2010). Previous empirical research has demonstrated a positive relationship between job stressors and job burnout (Jamal and Baba, 2000).

Further, enhanced job burnout impairs the creativity required in the big data context (Vecchio, 1990). We argue that burnout decreases creativity by occupying cognitive resources necessary for creativity. Specifically, job burnout, characterized by overwhelming exhaustion and feelings of frustration, cynicism, ineffectiveness and failure (Maslach and Jackson, 1981), consumes cognitive resources. In other words, job burnout leaves individuals with fewer resources to deal with tasks, in particular those requiring high creativity (Byron et al., 2010; Sacramento et al., 2013). As the cognitive resources necessary for creativity decrease, creativity is undermined (Byron et al., 2010). As an indirect support, prior research has shown that challenge stressors have negative, indirect relationships with performance through strains (Zhang, LePine, Buckman and Wei, 2014). Taken together, we thus propose the following hypothesis:

\[ H1a. \text{Job burnout mediates the negative relationship between challenge stressor and creativity.} \]

**Work engagement as a mediator**

Based on the JD–R model, Tetrick and Winslow (2015) posit that work engagement is a motivational process linking the stressors related to demands and recourses with organizational and individual outcomes. Work engagement is defined as a positive and fulfilling work-related state of mind characterized by vigor, dedication and absorption (Bakker et al., 2008; Schaufeli et al., 2002). Work-engaged employees have a “sense of energetic and effective connection with their work activities, and they see themselves as able to deal well with the demands of their jobs” (Schaufeli et al., 2006, p. 702).

First, we submit that employees faced with challenge stressors in a big data context are likely to enhance individual work engagement. As argued above, challenge stressors can promote mastery, personal growth or future gains, when they are appraised as related to learning and gaining confidence from personal experience (Cavanaugh et al., 2000). As mentioned by Crawford et al. (2010), although challenge stressors cause strains, challenge demand may also be positively related to engagement because they trigger positive emotions and active problem-focused coping styles that increase employees’ willingness to invest energy in efforts to meet this demand. Thus, higher levels of challenge stressors in a big data context would result in more enthusiasm, inspiration and a higher willingness to work.

Furthermore, challenge stressors may have a positive influence on creativity via the mediating role of work engagement. In their voluminous research on the antecedents of creativity, Zhou and George (2003) listed five routes through which individuals’ innate creativity can be awakened: identification, information gathering, idea generation, evaluation and modification and idea implementation. Engaged individuals might feel positive potential and take several strategies to solve the problem, such as defining the problem and then generating and evaluating alternative solutions (Lazarus and Folkman, 1984), rather than simply “fixing” problems (Tetrick and Winslow, 2015). As such, they are more likely to be dedicated to learning new skills to manage stressors. In addition, work engagement increases arousal, which elicits the use of creative thought and motivates persistence toward deriving solutions (e.g. Anderson et al., 2004), leading to enhanced creativity (Bunce and West, 1994). Therefore, we argue that work engagement mediates the relationship between challenge stressor and employee creativity:

\[ H1b. \text{Work engagement mediates the positive relationship between challenge stressor and creativity.} \]
The moderating role of CSE

Responses and appraisals to stressors vary as a function of individual characteristics, which “determines what is salient for well-being, shapes the person’s understanding of the event, and in consequence his or her emotions and coping efforts, and provides the basis for evaluating outcome” (Lazarus and Folkman, 1984, p. 55). Adopting the JD-R model, Tetrick and Winslow (2015) argue that CSE is a key personal resource, which influences stress management intervention. CSE refers to a basic, fundamental appraisal of one’s worthiness, effectiveness and capability as a person indicated by four well-established traits in personality literature: self-esteem, generalized self-efficacy, neuroticism and locus of control (Judge et al., 1997; Judge and Bono, 2001). In the current study, we posit the moderating effects of CSE on the effects of challenge stressors on work engagement and job burnout.

First, we propose that individual CSE attenuates the positive effect of challenge stressors on job burnout. There is a consensus in previous research that work stressors may be more threatening to individuals who are weak and vulnerable, psychologically or constitutionally or genetically (e.g. Lazarus and Folkman, 1984). Individuals with low levels of CSE might represent such prototypically weak and vulnerable people in the sense that they are much more emotionally reactive and view challenge stressors as difficult (Erez and Judge, 2001; Zhang, Kwan, Zhang and Wu, 2014). In addition, they feel lower in self-worth and capabilities (Chang et al., 2012). Low self-efficacy individuals doubt their own capability to perform and to cope successfully within an extensive range of situations (Chen et al., 2001). Previous research has indicated that generalized self-efficacy may influence the outcomes of stress, such that individuals with higher self-efficacy perform less poorly when facing heavy workloads and long working hours, compared with those with lower levels (Schaubroeck et al., 2000). In addition, individuals with lower self-esteem set a lower appraisal of their own self-worth (Rosenberg, 1965) and have less confidence in coping with hardship (Kammeyer-Mueller et al., 2009). As for locus of control, individuals with low CSE prefer to attribute their undesired effects to the powerful external environment rather than to their own control (Rotter, 1966). This perceived lack of control makes them desperate, so that higher levels of strain are generated by stressful situations (Spector et al., 2002). Combining these arguments, we propose:

\[ H2a. \] CSE attenuates the positive relationship between challenge stressor and job burnout such that the relationship is less positive when CSE is high.

Conversely, CSE is a key construct to facilitate the positive relationship between challenge stressor and work engagement. CSE is regarded as comprising an important personal resource that helps individuals to cope effectively with stressors by reinforcing positive appraisals of their jobs (Chang et al., 2012). Behaviorally, high-CSE individuals are likely to demonstrate high levels of job motivation and performance (Erez and Judge, 2001), with a tendency to seek active problem-solving skills and absorb new knowledge (Kammeyer-Mueller et al., 2009). Judge et al. (2000) demonstrated that individuals with higher levels of CSE not only perceive their job as more intrinsically satisfying, but are also more motivated to work actively and effectively, more willing to choose more complex jobs, and more engaged in work that gives them the opportunity to cope with difficulties. Hence, the higher the CSE level, the greater the chances individuals will face challenges with vigor, dedication and absorption. Thus, we posit that CSE strengthens the positive effect of challenge stressors on work engagement:

\[ H2b. \] CSE accentuates the positive relationship between challenge stressor and work engagement such that the relationship is more positive when CSE is high.
The moderating role of servant leadership

Employees’ interpretations of work stressors are also influenced by situational factors (Lazarus and Folkman, 1984), among which the leaders’ leadership style is one of the most important factors (Almatrooshi et al., 2016; Lepine et al., 2016; Zhang, LePine, Buckman and Wei, 2014). In the context of the management revolution, leaders play a key role in confronting challenges and then helping employees survive the transformation and achieving the benefits of that revolution (McAfee and Brynjolfsson, 2012). Servant leadership suggests that leaders lead by serving others (Greenleaf, 1970, 1977). They motivate and influence their followers by prioritizing followers’ interests over their own and trying to satisfy and fulfill followers’ personal and professional needs (Greenleaf, 1970; Liden et al., 2008; Van Dierendonck, 2011). In the current study, we explore the moderating effects of servant leadership on the relationship between challenge stressor on strain and work engagement.

We posit that servant leadership weakens the relationship between challenge stressor and job burnout. Under stressful work conditions, servant leaders are genuinely concerned with serving followers (Greenleaf, 1977; Gregory Stone et al., 2004) and provide immediate support so that followers can adapt to address stressors that come up at work (Van Dierendonck, 2011). They are sensitive to followers’ personal concerns and offer emotional healing and empathy to individuals faced with stressors (Spears, 1996). In addition, servant leaders highlight interpersonal relationship building, making a genuine effort to know, understand and support their followers (Ferris et al., 2009; Liden et al., 2008), which helps employees cope with challenge stressors more effectively. In addition, leadership competencies, such as cognitive, emotional and social intelligence competencies, contribute to the organization’s performance to cope with questions (Almatrooshi et al., 2016), which is also a key role of servant leadership. Overall, servant leaders provide employees with mental and physical resources, making employees less likely to feel psychologically strained when facing challenge stressors:

**H3a.** Servant leadership attenuates the positive relationship between challenge stressor and job burnout such that the relationship is less positive when servant leadership is high.

In contrast, we argue that servant leadership accentuates the positive effect of challenge stressor on work engagement. As servant leaders empower their followers by encouraging self-directed decision making and information sharing and coaching for innovative performance (Konczak et al., 2000), they foster a proactive and self-confident attitude among followers (Conger, 2000). In addition, they support and assist their followers by helping them solve complex tasks and face taxing job demands, which increases followers’ likelihood of overcoming challenges (Gregory Stone et al., 2004). Hence, servant leadership strengthens followers’ tendency to appraise challenge stressors as being consistent with their capability scope, making them experience higher levels of work engagement. Moreover, servant leadership also emphasizes individual growth and success, and demonstrates genuine concern for others’ career growth and development by providing support and mentoring (Luthans and Avolio, 2003). As a result, servant leadership is more likely to strengthen employees’ understanding of why they are being asked to face such challenge demands, and what potential future benefits can be expected in return, thus making them more likely to have a higher level of vigor and dedication to their work when confronted with challenge demands. Taken together, when employees, who are led by servant leaders, face challenge stressors, a positive, energetic and effective work-related state is expected. Thus, we hypothesize that servant leadership enhances the positive effect of challenge stressors on work engagement:

**H3b.** Servant leadership accentuates the positive relationship between challenge stressor and work engagement such that the relationship is more positive when servant leadership is high.
The integrative moderated mediation model

Thus far, we have developed theoretical underpinnings for the mediating effect of job burnout and work engagement as well as for the contingent effects of the CSE and servant leadership. The theoretical rationales behind the above hypotheses also suggest an integrative moderated mediation model.

Specifically, the theorizing behind $H1a$–$H2a$ and $H1b$–$H2b$ indicate that through augmenting or attenuating the association between challenge stressor and job burnout as well as work engagement, individual CSE affects the degree to which challenge stressors influence job burnout and work engagement, and subsequently employee creativity. Likewise, servant leadership, because of its moderating effect on the influence of challenge stressors on job burnout and work engagement ($H3a$–$H3b$), would influence the indirect effects of challenge stressors on creativity through job burnout and engagement ($H1a$ and $H1b$). Overall, we propose two sets of integrative moderated mediation hypotheses:

$H4a$. CSE attenuates the indirect negative effect of challenge stressor on creativity via job burnout such that the indirect negative effect decreases as CSE is higher.

$H4b$. CSE accentuates the indirect positive effect of challenge stressor on creativity via work engagement such that the indirect positive effect increases as CSE is higher.

$H5a$. Servant leadership attenuates the indirect negative effect of challenge stressor on creativity via job burnout such that the indirect negative effect decreases as servant leadership is higher.

$H5b$. Servant leadership accentuates the indirect positive effect of challenge stressors on creativity via work engagement such that the indirect positive effect increases as servant leadership is higher.

Methods

Participants and procedure

Data were collected from employees in six organizations located in China, covering both the manufacturing and service industries. The “big data driven” strategy was highlighted in their official documents. Indeed, in this management revolution era, both manufacturing and service organizations can potentially gain a competitive advantage as a result of the success of the big data driven strategy (McAfee and Brynjolfsson, 2012). As the power of big data is recognized, business leaders are increasingly pushing their organizations to implement the big data driven strategy (Gobble, 2013; Wamba et al., 2015).

To minimize common method bias, we used two sets of pencil-and-paper questionnaires to collect data from different sources: one from followers and the other from their immediate supervisors. Each follower was asked to complete a questionnaire that includes measures of challenge stressor, work engagement, job burnout, CSE and servant leadership, while each supervisor was asked to complete a questionnaire rating his or her followers’ creativity. Demographic information including gender and education were also collected from subordinates.

Measures

All measures in the study were originally in English. Applying the standard translation and back-translation procedure (Brislin, 1986), we translated the English scales into Chinese. Unless otherwise indicated, the measures were rated by the respondents on a five-point Likert-type scale. The anchors for the scale were from strongly disagree (1) to strongly agree (5).

Challenge stressor. A six-item scale developed by Zhang, LePine, Buckman and Wei (2014) was used to measure challenge stressor, which taps workload, time pressure,
task complexity, responsibility, and so on. The scale has been widely adopted in previous empirical research (e.g. Cavanaugh et al., 2000; LePine et al., 2004, 2005). One sample item is “I have to complete a lot of work.” Cronbach’s α was 0.87.

Work engagement. We measured work engagement using Schaufeli et al. (2006) shortened version scale, i.e., UWES-9. One sample item is “At work, I feel bursting with energy.” Cronbach’s α was 0.92.

Job burnout. Consistent with Zhang, LePine, Buckman and Wei (2014), we measured job burnout using a five-item emotional exhaustion scale from the Maslach Burnout Inventory (Maslach et al., 1996). One sample item is “I feel emotionally drained from work.” Cronbach’s α was 0.91.

Core self-evaluation. Judge et al.’s (2003) 12-item scale was used to measure the CSE. One sample item is “I am confident I get the success I deserve in life.” Cronbach’s α was 0.78.

Servant leadership. Servant leadership was measured using Liden et al.’s (2008) 28-item scale. This shortened scale has been adopted by Liden et al. (2014) and has been demonstrated to have good psychological properties. One sample item is “My leader puts my best interests ahead of his/her own.” Cronbach’s α for the scale was 0.89.

Creativity. Consistent with Yu and Frenkel (2013), we measured creativity using five items from the scale developed by Farmer et al. (2003). One sample item is “This employee seeks new ideas and ways to solve problems.” The Cronbach’s α was 0.93.

Control variables. We controlled for employees’ gender and education. Gender was coded as 1 = male and 2 = female. Education was coded into four categories (1 = high school or lower, 2 = associate’s degree, 3 = bachelor’s degree, 4 = master’s degree or higher).

Analysis strategy
We tested our hypotheses using path analytic procedures (Preacher et al., 2007) and conducted bootstrapping analysis to assess the significance of indirect effects (Shrout and Bolger, 2002). We used an SPSS macro (Hayes, 2012; Preacher et al., 2007) to estimate both mediation and moderated mediation models. Significance tests for indirect effects were based on bias-corrected confidence intervals derived from 5,000 bootstrapped samples (Preacher and Hayes, 2004a, b; Shrout and Bolger, 2002).

Results
Demographic summary
The final sample comprised 593 followers and 98 supervisors. Among the 593 followers, 251 were males (42.3 percent). In total, 33 of them had a high school or lower degree (5.6 percent), 167 an associate’s degree (28.2 percent), 326 a bachelor’s degree (55.0 percent) and 27 a master’s degree or higher (4.6 percent). Among the 98 supervisors, 52 were males (53.1 percent). In total, 33 of them had an associate’s degree (34.4 percent), 52 a bachelor’s degree (54.2 percent) and 8 a master’s degree or higher (8.3 percent).

Confirmatory factor analysis
Prior to hypotheses testing, we conducted a series of confirmatory factor analyses to examine the distinctiveness of the six studied variables (challenge stressor, CSE, servant leadership, job burnout, work engagement, creativity). As shown in Table I, the theorized six-factor model provided a good fit to the data ($\chi^2 = 2,712.62$, df = 845, RMSEA = 0.07, SRMR = 0.08, CFI = 0.93, TLI = 0.93) and showed a significantly better fit than the five-factor model ($\Delta \chi^2 = 1,869.11$, $p < 0.01$), the four-factor model ($\Delta \chi^2 = 3,358.08$, $p < 0.01$), the three-factor model ($\Delta \chi^2 = 3,768.47$, $p < 0.01$), the two-factor model
Given that the theorized six-factor model was superior in fit to all alternative models, we can continue to examine these variables as distinct constructs.

Descriptive statistics
The means, standard deviations and correlations among the variables are presented in Table II. As shown in the table, employees’ challenge stressor was positively related to job burnout \((r = 0.40, p < 0.01)\) but not related to work engagement \((r = 0.04, \text{ns})\). Additionally, job burnout was negatively related to creativity \((r = -0.07, p < 0.05)\), whereas work engagement was positively related to creativity \((r = 0.09, p < 0.05)\).

Hypotheses tests
H1a posits that supervisors’ job burnout mediates the relationship between employees’ challenge stressor and creativity. As indicated in M2 of Table III, there was a positive relationship between employees’ challenge stressor and their job burnout \((b = 0.50, p < 0.01)\) after controlling for employees’ gender and education. In addition, as shown in M14, employees’ job burnout was negatively related to their creativity after controlling for the demographic characteristics and challenge stressor \((b = -0.15, p < 0.01)\). Further, the results showed that the indirect effect of employees’ challenger stressor on their creativity

<table>
<thead>
<tr>
<th>Model</th>
<th>Factors</th>
<th>(\chi^2)</th>
<th>df</th>
<th>(\Delta\chi^2)</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>CFI</th>
<th>TLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>Theorized six factors</td>
<td>2,712.62</td>
<td>845</td>
<td>0.07</td>
<td>0.08</td>
<td>0.93</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>Five factors: job burnout and work engagement were merged as one factor</td>
<td>4,581.73</td>
<td>850</td>
<td>1,869.11**</td>
<td>0.11</td>
<td>0.11</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>Model 2</td>
<td>Four factors: challenge stressor, job burnout and work engagement were merged as one factor</td>
<td>6,070.70</td>
<td>854</td>
<td>3,358.08**</td>
<td>0.15</td>
<td>0.13</td>
<td>0.71</td>
<td>0.70</td>
</tr>
<tr>
<td>Model 3</td>
<td>Three factors: challenge stressor, job burnout, work engagement, and CSE were merged as a single factor</td>
<td>6,481.09</td>
<td>857</td>
<td>3,768.47**</td>
<td>0.15</td>
<td>0.13</td>
<td>0.69</td>
<td>0.67</td>
</tr>
<tr>
<td>Model 4</td>
<td>Two factors: challenge stressor, job burnout, work engagement, CSE and servant leadership were merged as a single factor</td>
<td>7,595.76</td>
<td>859</td>
<td>4,883.14**</td>
<td>0.16</td>
<td>0.14</td>
<td>0.63</td>
<td>0.61</td>
</tr>
<tr>
<td>Model 5</td>
<td>One factor: all variables were merged as a single factor</td>
<td>8,400.05</td>
<td>860</td>
<td>5,687.43**</td>
<td>0.17</td>
<td>0.14</td>
<td>0.59</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Notes: \(n = 593\). CFI, comparative fit index; RMSEA, root mean square error; SRMR, standardized residual mean root; NFI, normed fit index. \(*p < 0.05; **p < 0.01\)

**Table I.** Comparison of measurement models

**Table II.** Means, standard deviations and correlations among variables

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Gender</td>
<td>0.56</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Education</td>
<td>2.62</td>
<td>0.68</td>
<td>0.12**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Challenge stressor</td>
<td>3.66</td>
<td>0.72</td>
<td>0.01</td>
<td>0.23**</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Core self-evaluation</td>
<td>3.53</td>
<td>0.49</td>
<td>-0.09*</td>
<td>-0.09*</td>
<td>-0.11**</td>
<td>0.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Servant leadership</td>
<td>3.61</td>
<td>0.70</td>
<td>0.02</td>
<td>-0.19**</td>
<td>-0.12**</td>
<td>0.35**</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>6. Job burnout</td>
<td>2.41</td>
<td>0.90</td>
<td>-0.07</td>
<td>0.17**</td>
<td>0.40**</td>
<td>-0.48**</td>
<td>-0.31**</td>
<td>0.91</td>
</tr>
<tr>
<td>7. Work engagement</td>
<td>3.64</td>
<td>0.67</td>
<td>0.00</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.46**</td>
<td>0.50**</td>
<td>-0.31**</td>
</tr>
<tr>
<td>8. Creativity</td>
<td>3.30</td>
<td>0.90</td>
<td>-0.02</td>
<td>0.12**</td>
<td>0.10*</td>
<td>0.13**</td>
<td>0.07</td>
<td>-0.07*</td>
</tr>
</tbody>
</table>

Notes: \(n = 593\). Cronbach’s \(\alpha\) as the internal consistence reliabilities is on the diagonal. \(*p < 0.05; **p < 0.01\)
<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
<th>M8</th>
<th>M9</th>
<th>M10</th>
<th>M11</th>
<th>M12</th>
<th>M13</th>
<th>M14</th>
<th>M15</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>−0.16*</td>
<td>−0.14*</td>
<td>−0.20**</td>
<td>−0.18**</td>
<td>−0.13</td>
<td>−0.13</td>
<td>0.01</td>
<td>0.01</td>
<td>0.05</td>
<td>0.04</td>
<td>−0.02</td>
<td>−0.02</td>
<td>−0.04</td>
<td>−0.06</td>
<td>−0.04</td>
</tr>
<tr>
<td>Education</td>
<td>0.24**</td>
<td>0.12*</td>
<td>0.08</td>
<td>0.09</td>
<td>0.06</td>
<td>0.06</td>
<td>−0.04</td>
<td>−0.05</td>
<td>−0.03</td>
<td>−0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.16**</td>
<td>0.16**</td>
<td>0.15</td>
</tr>
<tr>
<td><strong>Independent variable</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Challenge stressor</td>
<td>0.50**</td>
<td>0.43**</td>
<td>0.45**</td>
<td>0.47**</td>
<td>0.47**</td>
<td>0.04</td>
<td>0.09*</td>
<td>0.08*</td>
<td>0.08*</td>
<td>0.08*</td>
<td>0.16**</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Moderators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Servant leadership</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Challenge stressor × CSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Challenge stressor × SL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mediators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job burnout</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work engagement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.04</td>
<td>0.18</td>
<td>0.38</td>
<td>0.39</td>
<td>0.24</td>
<td>0.25</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
<td>0.24</td>
<td>0.25</td>
<td>0.26</td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>$\Delta R^2$</td>
<td>0.14**</td>
<td>0.20**</td>
<td>0.01**</td>
<td>0.06**</td>
<td>0.01*</td>
<td>0.00</td>
<td>0.23**</td>
<td>0.01**</td>
<td>0.25**</td>
<td>0.01*</td>
<td>0.02**</td>
<td>0.01*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$</td>
<td>10.56**</td>
<td>40.79**</td>
<td>83.02**</td>
<td>69.47**</td>
<td>44.12**</td>
<td>36.52**</td>
<td>0.51</td>
<td>0.63</td>
<td>40.84**</td>
<td>34.97**</td>
<td>47.69**</td>
<td>39.20**</td>
<td>3.67**</td>
<td>4.69**</td>
<td>3.56**</td>
</tr>
</tbody>
</table>

**Notes:** $n=593$. CSE, core self-evaluation; SL, servant leadership. Unstandardized regression coefficients are shown. Adjusted $R^2$ are also shown. *$p < 0.05$; **$p < 0.01$.
via job burnout was significant (estimate = −0.06, 95% bootstrap CI [−0.14, −0.03]). Thus, H1a was supported.

H1b posits that employees’ work engagement mediates the relationship between employees’ challenge stressor and creativity. As indicated in M8 of Table III, after controlling for employees’ gender and education, the challenge stressor was not related to work engagement ($b = 0.04$, ns). Further, the results showed that the indirect effect of employees’ challenge stressor on creativity via work engagement was not significant (estimate = 0.01, 95% bootstrap CI [−0.01, 0.03]). Thus, H1b was not supported.

H2a suggests that employees’ CSE attenuates the positive relationship between challenge stressor and job burnout. As shown in M4 of Table III, after controlling for employees’ demographic characteristics, challenge stressor and CSE, the interaction term between challenge stressor and CSE was significant ($b = −0.09$, $p < 0.01$). Following the procedure recommended by Aiken and West (1991), we conducted simple slope tests. As indicated in Figure 2, for employees whose CSE was high, employees’ challenge stressor was less positively related to job burnout (simple slope = 0.33, $p < 0.01$) than employees whose CSE was low (simple slope = 0.57, $p < 0.01$). Thus, H2a was supported.

H2b suggests that employees’ CSE accentuates the positive relationship between challenge stressor and work engagement. As shown in M10 of Table III, after controlling for employees’ demographic characteristics, challenge stressor and CSE, the interaction term between challenge stressor and CSE was significant ($b = 0.07$, $p < 0.01$). As indicated in Figure 3, for employees whose CSE was high, challenge stressor was positively related to work engagement.
engagement (simple slope = 0.17, \( p < 0.01 \)), whereas for employees whose CSE was low, challenge stressor was not significantly related to work engagement (simple slope = −0.01, ns). Thus, \( H2b \) was supported.

\( H3a \) theorizes that servant leadership attenuates the positive relationship between challenge stressor and job burnout. As shown in M6 of Table III, after controlling for employees’ demographic characteristics, challenge stressor and servant leadership, the interaction term between challenge stressor and servant leadership was significant (\( b = −0.07, \ p < 0.05 \)). As indicated in Figure 4, when servant leadership was high, employees’ challenge stressor was less positively related to job burnout (simple slope = 0.37, \( p < 0.01 \)) than when servant leadership is low (simple slope = 0.57, \( p < 0.01 \)). Thus, \( H3a \) was supported.

\( H3b \) theorizes that servant leadership accentuates the positive relationship between challenge stressor and work engagement. As shown in M12 of Table III, after controlling for employees’ demographic characteristics, challenge stressor and servant leadership, the interaction term between challenge stressor and servant leadership was significant (\( b = 0.05, \ p < 0.05 \)). As indicated in Figure 5, when servant leadership was high, challenge stressor was positively related to work engagement (simple slope = 0.15, \( p < 0.01 \)), whereas when servant leadership was low, challenge stressor was not significantly related to work engagement (simple slope = 0.01, ns). Thus, \( H3b \) was supported.

To test \( H4a \) and \( H4b \), we examined the conditional indirect effects of employees’ challenge stressor on creativity through job burnout or work engagement at two values of
CSE (1 SD below the mean and 1 SD above the mean). As shown in Table IV, the indirect effect of challenge stressor through job burnout (diff (difference between conditional indirect effects) = 0.04, 95% bootstrap CI [0.01, 0.08]) or through work engagement (diff = 0.03, 95% bootstrap CI [0.004, 0.07]) differ significantly when employees’ CSE was at high vs low levels. Specifically, the indirect effect of challenge stressor through job burnout across low levels of CSE (estimate = −0.09, 95% bootstrap CI [−0.09, −0.02]) was stronger than that across high levels of CSE (estimate = −0.05, 95% bootstrap CI [−0.15, −0.03]). Thus, H4a was supported. Second, the indirect effect of challenge stressor through work engagement was significant across high levels of CSE (estimate = −0.18, 95% bootstrap CI [−0.33, −0.07]), but was not significant across low levels of CSE (estimate = −0.02, 95% bootstrap CI [−0.11, 0.07]). Thus, H4b was supported.

To test H5a and H5b, we examined the conditional indirect effects of employees’ challenge stressor on creativity through job burnout or work engagement at three values of servant leadership (1 SD below the mean and 1 SD above the mean). As shown in Table IV, the indirect effects of challenge stressor through job burnout (diff = −0.03, 95% bootstrap CI [0.001, 0.05]) differ significantly when servant leadership was at high vs at low levels. Specifically, the indirect effect of challenge stressor through job burnout at a low level of servant leadership (estimate = −0.09, 95% bootstrap CI [−0.15, −0.03]) was stronger than that at a high level of servant leadership (estimate = −0.06, 95% bootstrap CI [−0.11, −0.02]). Thus, H5a was supported. Second, the indirect effect of challenge stressor through work engagement was significant at a high level of servant leadership (estimate = 0.02, 95% bootstrap CI [0.004, 0.05]), but was not significant at a low level of servant leadership (estimate = 0.01, 95% bootstrap CI [−0.01, 0.03]). Thus, H5b was supported.

Discussion

Theoretical implications
First, our study contributes to the literature on talent management in a big data induced revolution context. As indicated by previous literature, organizations employing big data were more profitable, transparent, innovative and agile than their competitors (McAfee and Brynjolfsson, 2012; Wamba et al., 2015). Big data triggers not only technology but also managerial revolutions, among which the talent management revolution is a key factor. It is the talents who learn and use the big data and who should also be paid careful consideration in the managerial revolution. Specifically, our study simultaneously considers two key

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Mediator</th>
<th>Level of moderators</th>
<th>Indirect effect</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Challenge stressor</td>
<td>Moderator: CSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Job burnout</td>
<td>Higher</td>
<td>−0.05</td>
<td>[−0.15, −0.03]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower</td>
<td>−0.09</td>
<td>[−0.09, −0.02]</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td></td>
<td>0.04</td>
<td>[0.01, 0.08]</td>
</tr>
<tr>
<td></td>
<td>Work engagement</td>
<td>Higher</td>
<td>0.03</td>
<td>[−0.02, 0.02]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower</td>
<td>0.00</td>
<td>[0.01, 0.06]</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td></td>
<td>0.03</td>
<td>[0.004, 0.07]</td>
</tr>
<tr>
<td></td>
<td>Moderator: servant leadership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Job burnout</td>
<td>Higher</td>
<td>−0.06</td>
<td>[−0.11, −0.02]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower</td>
<td>−0.09</td>
<td>[−0.15, −0.03]</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td></td>
<td>−0.03</td>
<td>[0.001, 0.05]</td>
</tr>
<tr>
<td></td>
<td>Work engagement</td>
<td>Higher</td>
<td>0.02</td>
<td>[0.004, 0.05]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower</td>
<td>0.01</td>
<td>[−0.01, 0.03]</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td></td>
<td>0.01</td>
<td>[−0.01, 0.04]</td>
</tr>
</tbody>
</table>

Table IV. Results of moderated mediation effect
characteristics in the big data context: challenge stressor and creativity. Big data exert higher challenge stressors on employees, as employees must learn and use this new technology. For organizations to gain competitive advantages, a big data strategy also requires greater creativity from employees. This study explores the relationship between employees’ challenge stressors and creativity in the big data era, which adds new knowledge to the talent management literature.

As for our results, we proposed two opposing intermediate mechanisms to explain the effects of challenge stressors on employee creativity in the big data context, which can account for the inconsistent findings of challenge stressors’ effects on creativity. Consistent with previous studies, we found that challenge stressors exerted a negative effect on employee outcomes via the mediating role of job burnout (e.g., LePine et al., 2004; Podsakoff et al., 2007; Zhang, LePine, Buckman and Wei, 2014). In contrast, our findings regarding the intermediate effect of work engagement is, to some extent, counterintuitive. These results, together with those of previous studies, hint that the relationship between challenge stressor and work engagement might have boundary conditions.

Third, we investigated the dispositional and contextual factors that can accentuate the positive effect while attenuating the negative effect of challenge stressors on employee creativity. In the current study, we found that both CSE and servant leadership attenuated the harms of challenge stressors on creativity through job burnout but accentuated the benefits of challenge stressors through work engagement. On the basis of these findings, we extend the extant challenge stressor literature by identifying two important moderators and adding novel insights to the boundary conditions of its effects.

Managerial implications
The findings suggest several managerial and decision-making implications. For managers, this study highlights talent management in the big data era. Big data induce both technological and management revolutions. Talent management should garner more attention given this circumstance. Second, the level of challenge stressor and individual creativity appear as remarkable factors for individuals in the big data revolution, as the former is caused by the big data context and the latter is required by the big data context. Managers should protect challenge stressors from the influence of external strength and deal with them properly, as they will affect individual creativity. Third, managers can adjust the level of challenge stressors from work to employees to obtain better creative performance. Managers can expect a higher quality of creativity from those with a higher level of CSE under a similar level of challenge stressors in a big data context.

Employees in organizations facing increasing stressors brought about by the big data induced revolution must be mindful of the double-edged sword of challenge stressors. When they feel burned out and resource depleted, challenge stressors have a negative effect on their creativity. On the other hand, when they feel the support and services from their managers and organizations, challenge stressors can play a positive role on individual creativity via employees engaging in the work.

Limitations and future research
Our research is not without limitations. First, although this research extends our understanding of the big data era induced management revolution, from the perspective of talent management transformation, the relationship between challenge stressor and individual creativity and its underlying mechanisms investigated in the current research is only a small part of that transformation. As mentioned, as there are numerous challenges and revolutions in the big data era, many other issues are triggered and waiting to be explored. Specifically, as the characteristics of the management background are changing, previous research results may no longer hold and new issues may appear. In the context of a
new technology induced management revolution, more general research topics are expected. Second, although we have theoretically delineated the causal relationship among the variables, the cross-sectional data used in the current study still preclude causal conclusions. Future research should consider using multiphase data or a longitudinal research design to reexamine our research model. The last limitation concerns the potential generalizability of our research findings to other countries. In restricting our sample to Chinese organizations, the generalizability of our research findings to other countries might be limited.

We also suggest several future potential avenues. First, future research could assess the underlying mechanism that links challenge stressor with job burnout and work engagement. Second, our research only explores two boundary factors and mediators from challenge stressor to creativity. Future research could identify other boundary conditions. In addition to the research model studied in this paper, further research questions regarding the management revolution and talent management transformation would also be worth future investigation.

References


Further reading


Corresponding author
Xiao Jia can be contacted at: jiax.15@sem.tsinghua.edu.cn

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Innovating through digital revolution

The role of soft skills and Big Data in increasing firm performance

Francesco Caputo
Department of Pharmacy, University of Salerno, Fisciano, Italy

Valentina Cillo
Department for Research, Link Campus University, Rome, Italy

Elena Candelo
Department of Management, University of Turin, Turin, Italy, and

Yipeng Liu
Henley Business School, University of Reading, Henley-on-Thames, UK

Abstract

Purpose – The purpose of this paper is to investigate the relations among soft skill, information technologies and Big Data for building a possible bridge able to link human and technology dimensions for increasing firm performance.

Design/methodology/approach – Using the Business-focused Inventory of Personality, work personality of 4,758 human resources engaged in 72 high-tech European firms has been analyzed and its relations with firms’ investment in Big Data and firms’ economic performance have been tested using the structural equation modeling (SEM).

Findings – The research shows the existence of strong relations between some elements of human resources’ personality such as the work motivation and the social competencies and the firms’ economic performance. At the same time, the research clarifies the mediated effect of firms’ investment in Big Data in the relations between human resources’ organizational behavior and the firms’ economic performance.

Originality/value – The paper extends previous managerial contributions about Big Data management and human resource management providing evidence on which build more effective managerial models in the era of digital transformation.

Keywords Big Data, Artificial intelligence, Soft skills, High-tech European firms

Paper type Research paper

1. Introduction

For a long time, managerial and marketing studies have focused their attention on the role of “tangible resources” such as raw materials, production processes and physical structures as key levers for ensuring firms’ survival and competitiveness (Jauch and Glueck, 1988; Dakin, 1993). Nowadays, the challenging social and economic dynamics are showing the relevance of “not tangible resources” as key elements on which act for building managerial and business models able to support firms in catching emerging market opportunities (Teece, 1998; Chesbrough, 2010; Almatrooshi et al., 2016).

Among “not tangible resources,” an increasing relevance has been acquired by the multiple instruments rooted in the field of information and communication technologies (ICTs) (Roberts, 2000; Fuchs, 2009; Aquino et al., 2018; Gupta et al., 2018). The spread of social networks, virtual realities, electronic devices, 3D printers and artificial intelligence (AI) supported systems among the others are clear evidence of disruptive changes that are radically changing the world in which we all live.

Within the challenging debate about opportunities and risks provided by the ICT, several contributions have been specifically provided with reference to the multi-
trans-disciplinary domains of Big Data (McAfee et al., 2012; Chen et al., 2013) and AI (Liebowitz, 2001; Brodie and Mylopoulos, 2012). Accordingly, a number of theoretical perspectives have been adopted for examining firm innovation processes, including cognitive theory, dynamic capabilities theory, institutional theory, market orientation perspective, resource-based view and sociotechnical approaches (Huizingh, 2011; Garud et al., 2013).

According to the large part of these contributions, ICTs are considered as a source for process innovation, while process innovation is considered as catalyst for understanding the business value of ICT. In such a view, Big Data can support the application of business intelligence (BI) tools, while AI can support the definition of more efficient work processes (Chen et al., 2012).

Despite the multiple advancements in knowledge provided with reference to BI and AI, a large part of existing research approaches seems to be oriented to consider them as domains related only to firms’ processes and infrastructures (Chen and Zhang, 2014), while only few contributions have been provided with reference to the role of human resources in managing processes based on Big Data and AI (Powell and Dent-Micallef, 1997). Still, a general halo of uncertainty seems to linger with reference to the conditions that affect human resources in the adoption of Big Data and AI (Linoff and Berry, 2011).

With the aim to bridge this gap, the paper adopts an interpretative perspective centered on human resources for providing a possible conceptual framework able to link human and technology dimension under the shared umbrella of business management. The main aim of the paper is to identify possible link able to support an efficient and suitable combination between human resources’ features and opportunities offered by the advancements in knowledge in technology field. In such a line, the cognitive domain of 4,758 human resources engaged in 72 high-tech European firms is analyzed using the Business-focused Inventory of Personality (BIP) for defining key elements of human resources’ work personality. After this, the relations between key elements of human resources’ work personality, firm’s investment in Big Data and firms’ revenues have been analyzed using structural equation modeling (SEM) for tracing possible elements on which managerial researchers and practitioners can act for improving efficiency in the relation between firms’ human and technology dimension. Accordingly, the rest of paper is structured as follow: Section 2 defines the literature overview on which reflections herein are based and Section 3 proposes the hypotheses development. Section 4 illustrates methodology and research process. In Section 5, the results are presented and, in Section 6, they are discussed for underling theoretical and practical implications of the research. Finally, Section 7 proposes some final remarks, limitations and future directions for research.

2. Literature overview
The emerging trend toward open innovation requires an integrative perspective and it calls for a rethinking of traditional perspectives about firm boundaries for considering knowledge exploration, retention, and exploitation inside and outside organizational boundaries (Chesbrough, 2006; Del Giudice and Della Peruta, 2013).

Quantitative empirical studies about external knowledge sourcing provide evidence that involving a large number of external sources in innovation is a promising choice for improving firms’ economic performance (Chesbrough et al., 2006; Perkmann and Walsh, 2007; Del Giudice and Della Peruta, 2016). Open innovation scholars also agree that external sourcing of knowledge does not replace in-house research and development and they highlight the importance of “absorptive capacity” which allows firms to identify, absorb and make use of external knowledge (Spithoven et al., 2010).

Moreover, a large number of works have been concerned with the role of the diverse relationships and cooperation developed by firms with other stakeholders to “absorb”
The general idea is that the capacity of a firm to exploit external knowledge is a critical determinant of its capacity for innovation (Chesbrough, 2006; Scuotto, Del Giudice, della Peruta and Tarba, 2017).

As summarized in Table I, six knowledge capacities are needed to capture internal and external knowledge exploration, retention and exploitation: inventive, absorptive, transformative, connective, innovative and disruptive capacity (Argote et al., 2003; Lane et al., 2006).

While inventive capacity refers to internal exploring new knowledge, absorptive capacity relates to exploring external knowledge. Based on Cohen and Levinthal’s original definition of recognizing, assimilating and applying external knowledge, Zahra and George (2002) differentiated between potential and realized absorptive capacity. In a similar vein, Lane et al. (2006) distinguished exploratory, transformative and exploitative learning processes. Following this reconceptualization, the studies about the absorptive capacity in the knowledge management framework focus the attention on the knowledge acquisition (Gray, 2006).

From a different perspective, connectivity capacity is closely related to absorptive capacity. Accordingly, connective capacity comprises the process stages of maintaining knowledge in interorganizational relationships and subsequently reactivating this knowledge (Xia and Roper, 2008; Saviano and Caputo, 2013).

Studies about open innovation proposed by knowledge management literature suggest that firms should use external as well as internal ideas. Thus, new innovation models entail new forms of interactions and collaborations for fostering new products and processes development within varying contexts (Bellantuono et al., 2013; Saviano et al., 2018).

Generally speaking, innovation management process is a relevant part of the operations of many businesses (Slack et al., 2010; Scuotto, Santoro, Bresciani and Del Giudice, 2017). Innovation process management is a systematic approach for nurturing the creative capabilities of employees and for creating a workplace environment that encourages new ideas for workflows, methodologies, services and products (Ahmed and Shepherd, 2010).

Gartner’s recommendations to IT leaders interested in launching an innovation management program are to follow a disciplined approach based on five steps (Gartner, 1990): strategize and plan: settle on a shared view of aims and plans; develop governance: establish a process for making decisions; drive change management: build systems by which people can communicate and socialize via multiple channels; execute: make sure to draw from a wide range of sources to generate ideas for innovations that will transform the business, align the initiatives with business goals and then update and drive new elements of the initiatives in response to changing business requirements; and measure and improve: monitor and measure how the innovation process affects business outcomes.

Following Gartner’s recommendations, it is possible to note that digital transformation could be an interesting field for improving experience for both customers and employees. In the past, digital transformation was primarily integrated with business process management (BPM) tools, which aim to help companies in resources orchestration, routing work to the right people, manual task routine automation and self-service enabling where none existed before (Jeston, 2014). In a business landscape where companies need to be more agile,
BPM tools have helped line-of-business and IT departments to better align their goals and work processes (Smith and Fingar, 2003). Nowadays, new relevant opportunities for digital transformation are provided by the AI (Brodie and Mylopoulos, 2012) that combined with Big Data analytics (BBA) as “a collection of data and technology that accesses, integrates, and reports all available data by filtering, correlating, and reporting insights not attainable with past data technologies” (APICS, 2012, p. 6) are defining the “next management revolution”.

More specifically, Big Data has the potential to support firms in identifying opportunities related to decision-making processes and in defining more efficient organizational processes through the data acquisition, filtering and coding (Caputo, 2018). Big Data is a broad and abstract concept that is receiving great recognition both from scholars and practitioners (Caputo et al., 2018). Generally, it can be considered as a complex of tools for supporting firms’ decision-making process by using technology with the aim to rapidly analyze large amounts of variegated data (e.g., structured data from relational databases and unstructured data such as images, videos, e-mails, transaction data and social media interactions) from a variety of sources to produce a stream of actionable knowledge (Perko and Caputo, 2018).

Furthermore, following the reflections rooted in the field of Big Data, several contributions have underlined the support of Big Data for organization’s discovery of decision-making opportunities thanks to advanced analytics and data integration (Popović et al., 2012). According to Azma and Mostafapour (2012), there are two main features of data: the organizational learning process and the smart processing of data. The organizational learning includes the discovery of new knowledge and the dissemination of this knowledge; on the other hand, smart processing refers to the analysis and assessing of information with the aim to ensure the definition of efficient plans and the adoption of adequate control approaches.

From the process perspective, the main goal of the complex of technologies rooted in the field of Big Data is to improve the decision-making process reducing the time spent for the decision (Provost and Fawcett, 2013). From the product perspective, the domain of Big Data is considered as the complex of IT component that can be used to generate analytics for managers as the decision makers (Hazen et al., 2014). Finally, from the organizational perspective, Big Data field should be considered as a part of a decision environment that combines both technologies and human capacities for obtaining decisions aligned with firms’ plans (Chen et al., 2012). In nutshell, Big Data tools provide to the firms the opportunities for reducing the time needed for routinary processes and, in this way, they can support firms in paying more attention to the definition of vision and mission.

3. Hypotheses development

3.1 Knowledge management and work motivation

Recognizing the validity of previous contributions about knowledge management (Ruggles, 1998; Alavi and Leidner, 2001; Gold et al., 2001; Hussein et al., 2016), it is possible to state that knowledge management strongly affects the way in which firms define their relational approaches both from internal and external perceptive. With specific reference to the internal perspective, Ardichvili et al. (2006) show that firms’ attention to knowledge management influences the ways, in which human resources perceive their role in social and economic configurations contributing to the firms’ economic results. In the same directions, Jiang et al. (2012) underline that a strong firms’ attention on the levers of knowledge management influences the ways, in which human resources approach to their work in terms of motivation, engagement and commitment with several consequences in terms of their contributions to firms’ economic performance. Again, Yahya and Goh (2002) show that human resources’ work motivation is a consequence of the total amount of knowledge available inside the firms and it
is a consequence of the way in which firms manage the available knowledge. Finally, Al Mehrzi and Singh (2016) show that it exists a link between perceived organizational support and organizational culture mediated by employee motivation.

Recognizing the validity of all these contributions and considering firms’ revenues as a suitable manifestation of firms’ economic performance, it is possible to state that:

**H1.** There is a positive relationship between human resources’ work motivation and firms’ revenues.

### 3.2 Human resources’ social competence and digital transformation

Among the technology-based innovations, AI technology is becoming predominant in marketing fields such as customer relationship management (Ngai et al., 2009). By using supervised machine learning, BPM tools thanks to the support provided by the AI can support an easily identification of valuable targets in the market (Dumas et al., 2005). In such a vein, BPM tools and AI allow human employees to focus the attention on the more productivity processes (Weske et al., 2004).

In nutshell, managerial studies show that thanks to the combination between BPM and AI, human resources have the opportunities for better focusing their attention to the so-called human-based competencies (Wang and Wang, 2006). With regard to the complex of social competencies, human resource has the opportunities for better understanding markets’ expectations and needs with the aim to better delivering/providing firms’ services and products (Powell and Dent-Micallef, 1997). In this direction, Cappelli and Crocker-Hefter (1996) underline that human resources’ social competencies are a relevant lever on which firms should act for improving their market share and economic performance. In the same way, Huselid et al. (1997) underline that human resources’ social competencies direct impact on firms’ economic performance because they influence market’s perception about firms’ image. Again, Becker and Gerhart (1996) show that human resources with high social competencies provide a value added to firms interested in building loyalty-based relationships with the market for improving their economic performance. According to all these contributions and considering firms’ revenues as a suitable manifestation of firms’ economic performance, it is possible to state that:

**H2.** There is a positive relationship between human resources’ social competencies and firms’ revenues.

### 3.3 Big Data and human resources’ organizational behavior

Reflecting upon the general domain of Big Data, it is possible to state that after that data are collected and stored, the biggest challenge for firms is related to their analysis and to the extraction of valuable information (Carayannis et al., 2018). In such a perspective, Caputo et al. (2017) underline that Big Data and BBA can be efficiently used by firms only if human resources are able to effectively organize their work processes. More specifically, Brown et al. (2011) show that BBA act on the internal organizational flow and that the human resources’ organizational behaviors affect the opportunities for companies to adopt Big Data tools and approaches. In the same direction, LaValle et al. (2011) focus the attention on the way in which human resources affect the firm organizational behaviors stating that firms’ willingness to invest in Big Data depends by the human resources’ ability to understand and catch the opportunities provided by the so-called digital transformation. Accordingly, it is possible to speculate that:

**H3.** There is a positive relationship between human resources’ organizational behaviors and firms’ investment in Big Data.
Recognizing the validity of conceptualizations provided by managerial studies in the field of Big Data (Zikopoulos and Eaton, 2011; McAfee et al., 2012; Waller and Fawcett, 2013), it is also possible to state that an effective contribution of Big Data to firm’s organizations and decisions is possible only in the case in which human resources are able to understand the strategic value of Big Data (Carayannis et al., 2018). Using different words, as underlined by Bozionelos and Singh (2017), it is possible to state that human resources’ emotion intelligence (EI) affects the ways in which human resources perceive external world providing it a meaning and classifying through a subjective hierarchy of social and economic dynamic. In such a vein, Amendola et al. (2018) underline that human resources perspectives (or more generally emotions) define the firms’ approach toward innovation affecting their willingness and orientation in the adoption of new technologies. In the same direction, Pradhan et al. (2017) demonstrate that human resources’ EI influences employees’ adaptation and then the possibility for success of firms’ strategies. From a different perspective, Sagiroglu and Sinanc (2013) show that only firms with human resources inspired by an innovative conceptualization of the market are oriented to start new (also technology-based) processes. Following these contributions, the paper states that:

H4. There is a positive relationship between human resources’ emotions and firms’ investment in Big Data.

Furthermore, reflecting upon the role of Big Data in firms’ managing processes, several contributions should be considered also with reference to the impact of Big Data on firms’ economic performance (Chae et al., 2014; Wamba et al., 2017). In such a view, Wamba et al. (2017) demonstrate that firms’ attention to Big Data, and more generally, to ICT positively impacts on firms’ economic performance, while Brown et al. (2011) empirically verify the link between firm’s attention and investment in Big Data and firms’ market performance. From a different perspective, Weill (1992) shows that Big Data improves the quality of several firms’ internal processes with positive effects on firms’ economic performance, while Provost and Fawcett (2013) show the high contribution provided by Big Data to firms’ selling strategies and the related positive impact of firms’ revenues. According to all these contributions, the paper speculates that:

H5. There is a positive relationship between firms’ investment in Big Data and firms’ revenues.

4. Methodology and research process
4.1 Sample and data collection
Considering the pervasive nature of digital transformation for European companies involved in high-tech sectors (European Commission, 2017), a sample of 1,175 human resources engaged in 23 high-tech European firms is analyzed using a questionnaire based on the BIP (Hossiep and Paschen, 2008). BIP is a psychological questionnaire used for investigating professional characteristics of human resources (Birknerova, 2012). The questionnaire is composed by 210 through which human resources’ occupational orientation and occupational behavior are measured via a six-point-based scale (Kauer et al., 2007; Langendörfer, 2008).

The choice to focus the attention on the high-tech firms is motivated by their strong attention on technology-based innovations and to the opportunities for better identifying their investment in the field of Big Data using official websites and documents (CIONET and Next Value, 2017).

Building upon the “SEP ELITE Tech Scaleup 100” launched by Mind the Bridge in partnership with ELITE, London Stock Exchange Group’s business support and capital raising program (Mind the Bridge and ELITE, 2018), all the firms included in the list have been contacted via e-mail explaining the aim of the research and describing the
research process. After the first contact, 86 firms have declared their interest in participating in the research and a link to a questionnaire composed by 70 items inspired by BIP model has been sent to the firms that have declared their interest in participating to the research requiring them to share the link with their human resources. Therefore, the research is based on a survey methodology, which is useful to enhance the generalization of results.

The questionnaire was directed to investigate human resources’ perceptions about four independent variables: human resources’ work motivation, human resources’ social competencies, human resources’ organizational behavior and human resources’ emotions management. According to the BIP model, these categories have been split up in items and, for each items, a certain number of sentences have been formulated requiring to the firms’ human resources to express their opinion using a seven-point Likert scale in which 1 means “strongly disagree” and 7 means “strongly agree.”

In Table II, they are summarized the main contents of the questionnaire.

During the data collection period, 14 firms changed their opinion declaring that it was not possible to collect data about the perceptions of their human resources so the final sample is composed by 72 high-tech European firms. From these firms, 4,758 human resources have completed the questionnaire from June to November 2018.

During the research, data related to the companies inside the sample such as firms’ revenues and firms’ investment in Big Data projects in the period from 2015 to 2017 have been collected using official data set provided by Amadeus (2018) – a Bureau van Dijk data set (https://amadeus.bvdinfo.com/).

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Items</th>
<th>Items’ meaning</th>
<th>Number of sentences used for measuring the item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human resources’ work motivation</td>
<td>Results orientation</td>
<td>Capability to work for producing results</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Power orientation</td>
<td>Capability to motivate other human resources using recognized structures and processes</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Leadership orientation</td>
<td>Capability to motivate other human resources using social influence</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Feeling</td>
<td>Capability to understand relational signals</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Capabilities for building</td>
<td>Capability for building relational network and long-term interactions</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>relationships</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sociality</td>
<td>Capability for building conditions for collaboration and positive interactions</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Team orientation</td>
<td>Capability to work in team and to support team work</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Assertiveness</td>
<td>Capability to defend personal idea and identity</td>
<td>5</td>
</tr>
<tr>
<td>Human resources’ social competencies</td>
<td>Conscientiousness</td>
<td>Capability to perform the work according to shared standards and procedures</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Flexibility</td>
<td>Capability to adapt himself to external changes</td>
<td>5</td>
</tr>
<tr>
<td>Human resources’ organizational behavior</td>
<td>Action orientation</td>
<td>Capability to product results</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Emotion stability</td>
<td>Capability to manage personal emotions</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Capacities to work under</td>
<td>Capability to produce results also in condition of physical and/or psychological stress</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>pressure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human resources’ emotions management</td>
<td>Self-loyalty</td>
<td>Capability to maintain personal position as consequence of the emotional independence</td>
<td>5</td>
</tr>
</tbody>
</table>

*Table II. Variables under investigation, meanings and number of items*
4.2 Data analysis

The data collected have been analyzed respecting the rules about the privacy and adopting all the techniques for avoiding possible biases (Ott and Longnecker, 2015). The data have been analyzed and organized with the aim to test the hypotheses reported in Table III.

The hypotheses have been tested via covariance-based SEM using IBM® SPSS® Statistics – Version 25. According to Ullman and Bentler (2012), SEM is “a collection of statistical techniques that allow a set of relationships between one or more independent variables (IVs), either continuous or discrete, and one or more dependent variables (DV), either continuous or discrete, to be examined” (p. 661). The research approach is based on SEM because it is a method through which it is possible “over performing a series of multiple regressions; namely, it provides a test of the overall model fit” (Savalei and Bentler, 2006, p. 339). Following Reinartz et al.’s (2009) suggestion, a covariance-based SEM has been conducted because it outperforms variance-based SEM in terms of parameter consistency and is preferable in terms of parameter accuracy.

Before conducting the test via SEM, the relations among the variable were examined for analyzing common method bias (MacCallum and Austin, 2000; Bandalos, 2002). Harman’s single factor test to extract a single factor indicated an explained variance of 27.93 percent, providing an evidence about the absence of common method bias (Malhotra et al., 2006). According to Podsakoff et al. (2003), “the basic assumption of this technique is that if a substantial amount of common method variance is present, either (a) a single factor will emerge from the factor analysis or (b) one general factor will account for the majority of the covariance among the measure” (p. 889).

Finally, according to Lomax and Schumacker (2004), $R^2$ was measured for defining the capability of dependent variables to explain the variance of independent variables. The research shows that the measurement model has an $R^2$ equal to 0.75 evidencing a strong effect size (Moore et al., 2013).

Building upon the hypotheses summarized in Table III, the conceptual model of the research has been defined as reported in Figure 1.

To evaluate the model fit, a number of incremental, absolute and parsimonious fit indices were measured (Barrett, 2007; Steiger, 2007), including the goodness of fit index (GFI), the normed fit index (NFI), the comparative fit index (CFI), the standardized root mean square residual (SRMSR) and the root mean square error of approximation (RMSEA).

5. Findings

5.1 Internal consistency reliability and construct validity

For verifying the reliability of the data collected through the questionnaire, the Cronbach’s $\alpha$ coefficients were measured with reference to all the independent variables. A Cronbach’s $\alpha$ value equal or higher than 0.7 is considered suitable for applied research (Nunnally, 1978), while a Cronbach’s $\alpha$ value equal or higher than 0.6 can be considered suitable in the case of exploratory research (Hair et al., 2012). As reported in Table IV, all Cronbach’s $\alpha$ coefficients exceed the cut-off value of 0.7.

<table>
<thead>
<tr>
<th>Hypothesis (H)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 (+)</td>
<td>There is a positive relationship between human resources’ work motivation and firms’ revenues</td>
</tr>
<tr>
<td>H2 (+)</td>
<td>There is a positive relationship between human resources’ social competencies and firms’ revenues</td>
</tr>
<tr>
<td>H3 (+)</td>
<td>There is a positive relationship between human resources’ organizational behaviors and firms’ investment in Big Data</td>
</tr>
<tr>
<td>H4 (+)</td>
<td>There is a positive relationship between human resources’ emotions and firms’ investment in Big Data</td>
</tr>
<tr>
<td>H5 (+)</td>
<td>There is a positive relationship between human resources’ organizational behaviors and firms’ revenues mediated by firms’ investment in Big Data</td>
</tr>
</tbody>
</table>

Table III. Hypotheses under investigation
At the same time, the construct validity was analyzed combining convergent validity and discriminant validity. Specifically, convergent validity was measured by calculating the average variance extracted (AVE), while discriminant validity was verified by comparing the square roots of AVEs to the correlations between constructs. As reported in Table IV, the square roots of AVEs were all greater than their respective relationships, providing solid evidence of discriminant validity.

5.2 Hypothesis testing via SEM
Via SEM using IBM® SPSS® Statistics – Version 25, the hypotheses were tested and the results are reported in Table V. According to Hooper et al. (2008), all the hypotheses with a p-value lower than 0.05 can be considered verified with reference to the analyzed sample.

Finally, for verifying the model fit, several statistics such as GFI, NFI, CFI, SRMSR and RMSEA were measured. As reported in Table VI, all the cut-off values are exceeded.
6. Discussions

According to the results reported in Table V, the study demonstrates the existence of positive relationships between human resources’ work motivation and firms’ revenues (H1). This result is aligned with previous studies provided by the research works rooted in the field of human resource management (HRM) (Schuler and MacMillan, 1984; Schneider, 1988; Ulrich et al., 1995). Specifically, as underlined by Wright et al. (1994), the human resources’ abilities to achieve the planned aims, organize the flow processes and stimulate collaborative relationships are one of the key levers for firms’ competition. In such a vein, Barney and Wright (1998) state that human resources are the first source of firms’ competitive advantages and Lengnick-Hall and Lengnick-Hall (1988) underline that human resources able to be aligned with firms’ plans and are also able to stimulate firms’ in achieving greater and more ambitious results. From a different perspective, Li et al. (2006) underline that human resources’ work motivation is one of the emerging topics on which business researchers and practitioners should reflect for improving firms’ performances and Putra et al. (2017) demonstrate that firms endowed by human resources strongly motivated are able to better catch market opportunities and face market challenges.

The results also show that there is a positive relationship between human resources’ social competencies and firms’ revenues (H2). This result can be considered a partial demonstration of previous sociological studies about the high relevance of social interactions within firms’ environment as a way for improving firms’ efficiency (Gamst, 1991). As underlined by Caputo and Evangelista (2019), human resources spend a large part of their life time in work environment and, for this reason, they demand to satisfy their need for social interactions as basic conditions for ensuring their productivity. From a different perspective, Cabrera and Cabrera demonstrate that social interactions within work environment are an efficient way for ensuring informal information flows through which human resources have the opportunities for aligning their behaviors to firms’ plans supporting the achievement of firms’ aims.

Furthermore, the study also demonstrates the positive relations between human resources’ organizational behaviors and firms’ investment in Big Data (H3). This result can be considered aligned with previous contributions provided with reference to the conditions required for a useful use of Big Data inside firms’ processes (Leefflang et al., 2014). As clarified by Han et al. (2011), Big Data is a complex set of tools that improve human resources’ ability to read available data about specific dynamics and domains. Accordingly, Big Data orientation can

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 (+): HRs’ work motivation → firms’ revenues</td>
<td>0.0412</td>
</tr>
<tr>
<td>H2 (+): HRs’ social competencies → firms’ revenues</td>
<td>0.0238</td>
</tr>
<tr>
<td>H3 (+): HRs’ organizational behaviors → firms’ investment in Big Data</td>
<td>0.0481</td>
</tr>
<tr>
<td>H4 (+): HRs’ emotions → firms’ investment in Big Data</td>
<td>0.0760</td>
</tr>
<tr>
<td>H5 (+): Firms’ investment in Big Data → firms’ revenues</td>
<td>0.0437</td>
</tr>
</tbody>
</table>

Table V. Results of hypothesis testing via SEM

<table>
<thead>
<tr>
<th>Model fit index</th>
<th>Cut-off values (source)</th>
<th>Measured value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFI</td>
<td>&gt; 0.90 (Jöreskog and Sörbom, 1996)</td>
<td>0.921</td>
</tr>
<tr>
<td>NFI</td>
<td>&gt; 0.90 (Hu and Bentler, 1999)</td>
<td>0.903</td>
</tr>
<tr>
<td>CFI</td>
<td>&gt; 0.90</td>
<td>0.971</td>
</tr>
<tr>
<td>RMSEA</td>
<td>&lt; 0.06 (Hooper et al., 2008)</td>
<td>0.031</td>
</tr>
<tr>
<td>SRMR</td>
<td>&lt; 0.08 (Hooper et al., 2008)</td>
<td>0.035</td>
</tr>
</tbody>
</table>

Table VI. Model fit statistics
emerge only in the case in which firms are equipped with human resources able to understand the effective role of Big Data and digital transformation (Loebbecke and Picot, 2015). With reference to this, Sagiroglu and Sinanc (2013) show that Big Data can become a problem for firms in the case in which they are not approached in the correct way, while Carayannis et al. (2018) show that only acting on human resources’ soft skills Big Data can support firms’ processes through the definition of so-called Wise Data.

Finally, the research demonstrates that there is a positive relationship between firms’ investment in Big Data and firms’ revenues ($H_5$). This result is aligned with previous studies about the managerial contributions of Big Data (Prescott, 2014; Frisk and Bannister, 2017; Dubey et al., 2018). Specifically, Wamba et al. (2017) have demonstrated the positive impact of Big Data use on firms’ performance, while Brown et al. (2011) have demonstrated the multiple advantages for firms related to an adequate investment and a correct use of Big Data.

Despite the general validity of the research herein not all the hypotheses are verified. Specifically, the study does not provide validation about the existence of positive relationships between human resources’ emotions and firms’ investment in Big Data ($H_4$). This result can be discussed in the light of the complexity related to the domain of human emotions (Lane and Nadel, 1999). As clarified by Frey and Stutzer (2010), humans’ emotion impacts all social and economic processes but the real challenge is to identify the correct way for measuring this impact. In the same direction, Powell and Dent-Micallef (1997) reflect about the role of human resources’ emotions on firms’ orientation toward innovation, and they recognized the complexity for defining instruments able to support human resources in understanding their emotions and communicating them in a suitable way.

### 6.1 Theoretical implications

According to the results reported in previous sections, it clearly emerges the need for extending the boundaries of social and managerial studies interesting in business performances with the aim to clarify the elements able to influence on work motivations. In such a perspective, managerial and marketing researchers should develop multi- and trans-disciplinary research paths aimed at defining approaches through which stimulate human resources involvement, engagement and commitment in firms’ strategies and actions (Geroy et al., 2000; Wright et al., 2000).

The research also shows the need for developing theoretical framework able to explain paths and implications of social interactions in work environment as a way for increasing information sharing and employees’ commitment (Chrusciel, 2006; Mittal and Dhar, 2015).

Again, discussed results underline the need for increasing the attention on human resources’ ability to organize work processes and activities (Tikkanen et al., 2005; Kesting and Parm Ulhøi, 2010). In such a perspective, the research calls the attention of managerial researchers on the need for developing instruments and processes through which increase not only human resources’ problem-solving competences but also their decision-making competencies (Vaiman et al., 2012; Alpkann et al., 2010).

Finally, the research enforces the relevance of ongoing debate about the role of Big Data in managerial fields underling the need for in-depth investigation in which ways human resources and firms’ processes can be modified for catch all the opportunities offered by the digital transformation (Liu et al., 2011).

### 6.2 Practical implications

From practical point of view, the research underlines the need for developing tangible instruments through which stimulate human resources’ work motivation. In such a perspective, it emerges the need for defining paths for human resources engagement based on individual personality and cognitive domain (Hess and Bacigalupo, 2011; Wang et al., 2013).
A possible advancement for business practices with reference to this point could be based on the firms’ adoption of psychological instruments able to provide useful information about human resources’ personality on which build personalized techniques for engagement and motivation (Canós-Darós, 2013).

According to the results of the study proposed in previous sections, it also emerges the need for managerial practitioners to develop approaches able to stimulate social interactions within work environment (Andriopoulos, 2001). In such a way, differently from the consolidated approaches based on the maximization of time spend in work processes, the research underlines the need for increasing time for social interactions as a way for improving firms’ performance (Lin and Lee, 2006).

Again, the research calls the attention of managerial practitioners on the need for developing education processes not only based on the enforcement of human resources technical skills but also oriented to increase human resources’ ability in planning their activities and work processes (Luoma, 2000).

Finally, the research demonstrates the validity of business approaches based on the investment in Big Data (Zeng and Khan, 2018). From this point of view, it emerges the need for managerial practitioners to accurately define the processes through which face the ongoing digital transformation for firms identifying the correct way for balancing human and technology dimensions (Dawes and Rowley, 1998).

7. Final remarks, limitations and future directions for research

The general domain of digital transformation and the related research field interested in Big Data, BPM and AI can be considered the source of one of the most challenging debate in the field of managerial studies (Matt et al., 2015). Over the last few years, an increasing number of researchers and practitioners have provided managerial studies about the impacts of digital transformation on firms’ organizations and processes (Zhu et al., 2006). At the state, the general part of these contributions focuses the attention on the digital (or technology) dimensions of the opportunities and processes that are emerging from the digital transformation without consider their impact and/or relations with the human dimensions of the firms (Perko and Caputo, 2018).

With the aim to bridge this gap, the paper tries to build a conceptual framework direct to link soft skills of human resources and the firms’ orientation to technology-based innovations. The final aim is to partially clarify the relations between human and technology dimensions inside firms’ environment aftermath the so-called digital transformation.

Accordingly, the research demonstrates the existence of several connections between human and technology dimensions in firms’ setting opening to multiple managerial implications. In such a vein, as a result of the research herein it is possible to state that companies interesting in catching the opportunities provided by the digital transformation needed to act on human resource management and to build the conditions for an effective use of the multiple emerging instruments related to digital transformation inside the firm. From this perspective, several general recommendations can be derived as following summarized:

- Prioritize hybrid approaches: initially, technology-based tools should be leveraged toward only repeated, low-skill task. More complex tasks, like the use of AI for automatizing the supply chain it is needed wait until human resources do not develop an adequate level of confidence with the underlying technology.

- Identify feasibility across the firm: technology-based tools are most useful if they are adopted in firms’ processes for which high skills of human resources are required. In these processes, technology-based tools can increase the productivity of firms building conditions for firms’ competitive advantage.
Align data with goals: firms should focus their attention on the collection and analysis of data related to their business activities. In such a perspective, firms should understand that not all the data are useful but that data can contribute to firms’ performance only if they are analyzed in the light of efficient interpretation schemes for creating condition of wisdom.

Look for “semi-structured data”: analytics solutions are useful in finding trends in either structured data but semi-structured data could offer several opportunities for matching human creativity with technology advantages.

Summarizing the reflections and empirical results herein, the paper calls for the attention on the multiple dimensions of so-called digital transformation and specifically on the underestimated relevance of human resources in the firms’ transformation and processes (Bowersox et al., 2005). Accordingly, the role of human resources is analyzed using a holistic approach with the aim to link in a common conceptual framework several relevant elements of ongoing digital transformation.

Considering the pervasive nature of the digital revolution (Dreyer et al., 2006), the reflections and empirical results herein can be considered only a small part of a more widen conceptual framework to which both researchers and practitioners are called to contribute for clarifying the pillars that will affect the next changes of the world in which we all live.

References


Further reading


Corresponding author
Francesco Caputo can be contacted at: fcaputo@unisa.it

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Big data and dynamic capabilities: a bibliometric analysis and systematic literature review

Riccardo Rialti
DISEI – Dipartimento di Scienze per l’Economia e l’Impresa,
Università degli Studi di Firenze, Firenze, Italy

Giacomo Marzi
Lincoln International Business School, Department of Strategy and Enterprise,
University of Lincoln, Lincoln, UK

Cristiano Ciappei
DISEI – Dipartimento di Scienze per l’Economia e l’Impresa,
Università degli Studi di Firenze, Firenze, Italy, and

Donatella Busso
Dipartimento di Management, Università degli Studi di Torino, Torino, Italy

Abstract
Purpose – Recently, several manuscripts about the effects of big data on organizations used dynamic capabilities as their main theoretical approach. However, these manuscripts still lack systematization. Consequently, the purpose of this paper is to systematize the literature on big data and dynamic capabilities.

Design/methodology/approach – A bibliometric analysis was performed on 170 manuscripts extracted from the Clarivate Analytics Web of Science Core Collection database. The bibliometric analysis was integrated with a literature review.

Findings – The bibliometric analysis revealed four clusters of papers on big data and dynamic capabilities: big data and supply chain management, knowledge management, decision making, business process management and big data analytics. The systematic literature review helped to clarify each clusters' content.

Originality/value – To the authors' best knowledge, minimal attention has been paid to systematizing the literature on big data and dynamic capabilities.

Keywords Performance, Systematic literature review, Big data analytics, Dynamic capabilities, Bibliometric analysis, Big data

Paper type Research paper

1. Introduction

Big data has dramatically affected the traditional ways of running a business in the twenty-first century (Chen et al., 2012; McAfee and Brynjolfsson, 2012). In the current technological era, big data managers have an almost infinite amount of detailed information at their disposal (Erevelles et al., 2016). Therefore, it is widely assumed that big data will allow managers to be increasingly informed on the state of internal operations, supply chain processes, workforce’s performances and the consumers’ behavioral patterns (Bresciani et al., 2017). Management decision-making processes are simultaneously evolved and, as such, managers get the ability to decide upon the most suitable strategy to implement according to the newly available information (Chen et al., 2012).

In the light of the emerging potential of big data, it is evident that there is a requirement of studies about such systems and related organizational capabilities. It is needed to decodify and transform these data sets into insights, management decisions and organizational performances (Labrinidis and Jagadish, 2012). Thus, scholars started to point out how big data are such complex data sets, formed by heterogeneous data, that may be analyzed only using big data analytics (BDA) systems (Rialti et al., 2018). Only information
systems capable of processing different kind of data formats simultaneously may ensure that the necessary information flows in big data era (Vera-Baquero et al., 2016). On the other hand, as organization-wide capabilities are necessary to make everyone use complex technological systems, BDA capabilities have started receiving attention simultaneously from pertinent literature (Chen et al., 2012). It has emerged how organizations should foster the development of internal technical, managerial and personnel capabilities in order to exploit the big data and properly use BDA systems (Akter et al., 2016).

In case an organization is capable of implementing proper systems and develop the right capabilities, the true potential of big data availability may then emerge. Accordingly, big data were seen to be linked to increased organizational performances in terms of agility, flexibility and ambidexterity (Rialti et al., 2018). With this, the apparent impact of big data on dynamism started. Specifically, an organization may be able to scan the environment constantly and obtain a competitive edge with such capabilities. Consequently, big data, BDA systems and BDA capabilities were observed to influence value creation processes (Wamba et al., 2017).

Several different streams from different literatures have explored how organizations can exploit big data to create value. Researchers dealing with informatics and information systems were the first to explore the importance of big data (Chen et al., 2014). Obviously, theoretical approaches used by different scholars were related to the streams of literatures their papers were focused on. Appropriately, in terms of managerial literature, scholars have mostly used dynamic capabilities – even if sometimes in conjunction with other theories like IT business value or knowledge-based view – as the principal theoretical approach to understand how big data are affecting organizations (Côrte-Real et al., 2017).

Dynamic capabilities represent a suitable approach to study the effect of information systems or their specific capabilities on the organizations (Contractor et al., 2016). The utilization of BDA capable systems is frequently linked to common processes and routines that may be used to solve different data-related problems (Wamba and Mishra, 2017). The adaptable BDA systems are usable in different situations and may provide a competitive edge during environmental turbulence (Akter et al., 2016). Similarly, BDA capabilities are a set of capabilities that may help an organization to adapt an existing resource base (in this case data) to address different information needs emerging in different situation (Rialti et al., 2018). As both these considerations are coherent with dynamic capabilities theory, it emerges that dynamic capabilities are the most used approach in research about big data and performance (Wamba et al., 2017).

In spite of this growing interest in big data, dynamic capabilities and organizational performances, these manuscripts lack a proper systematization. Consequently, it is clear that there is necessity of mapping and systematizing existing literature (Braganza et al., 2017). To do this, the authors have performed a bibliometric analysis to map the knowledge concerning this stream of research and they have systematically reviewed literature to explore content of the most relevant papers (Caputo et al., 2018).

This paper is structured as follows: the following paragraph analyses the importance of big data and BDA systems and their capabilities to organizations, and the contributions of dynamic capabilities in this stream of literature; next, the methodological procedure is described; then, the results of the bibliometric analysis and the systematic literature review are presented. Finally, the authors present their suggestions for future research.

2. Theoretical background

2.1 Big data: the revolution has arrived

According to the seminal research by McAfee and Brynjolfsson (2012), “smart leaders across industries will see using big data for what it is: a management revolution” (p. 5). Several years later, the magnitude of the impact of big data across the management world is
clearly visible to everyone involved with big data. While information has always been identified as one of the most important value-creating factors for any organization, big data characteristics have brought information value-creation potential to an unprecedented level (El-Kassar and Singh, 2018).

Big data differs from traditional large data sets in terms of its volume, velocity, variety, veracity, value, variability and visualization – a.k.a. “Seven Vs of Big Data” (Mishra et al., 2017). Volume refers to the sheer dimensions of the typology of data sets (McAfee and Brynjolfsson, 2012). Indeed, big data’s dimensions frequently exceed the terabyte (Mishra et al., 2017). Velocity of big data is the “rate at which data are generated and the speed at which it should be analysed and acted upon” (Gandomi and Haider, 2015, p. 138). Variety is related to the “heterogeneous sources of big data (i.e. sensors embedded in machines, consumers’ activities on social media, B2C or B2B digital interactions, etc.) and the consequent heterogeneous formats that the files composing big data may assume” (Rialti et al., 2018, p. 7). Veracity is related to the necessary degree of trustworthiness that the sources of big data must possess (Mishra et al., 2017). Value is linked to the economic value that may be generated by an organization due to processes and technologies that analyze big data (Xuemie, 2017). Variability refers to the possible variations in data flow rate, processing and data sources (Wamba and Mishra, 2017). Finally, visualization concerns the possibility for data analysts to get visual insights as an output of big data analysis (Mishra et al., 2017).

As it is deducible from big data’s characteristics, the extraction of insights from these data sets poses unprecedented challenges to organizations. Big data are indeed so large and complex data sets that cannot be processed using traditional database software (Labrinidis and Jagadish, 2012). Data lakes, NoSQL data models, schema-less data retrieval, machine learning and other tools based on artificial intelligence paradigms are necessary to collect, store and analyze big data. As such, organizations may define ad-hoc BDA processes at the following stages: data acquisition, cleansing, integration, modeling and interpretation (Labrinidis and Jagadish, 2012).

Coherently, organizations should focus on developing BDA systems capable of supporting such processes. Thus, BDA systems should not only be capable of collecting data, but also clearing it from unworthy components (i.e. spam messages or messages without any useful content), modeling data, and obtaining information that could generate competitive advantages and economic value (Prescott, 2014). The implementation of these systems is not without their challenges. First of all, BDA systems usually need extremely large networked hardware's architecture, and need to rely on cloud storage and computing, and require extremely fast internet connections (Gandomi and Haider, 2015). Infrastructural and technological complexity is the first problem. Second, very frequently, such systems are built around such complex infrastructures and architectures, or depend on extremely complicated computer-science analytics methods, that managers and employees may reject the implementation of these systems (Xuemie, 2017), as they may not understand how such systems work. Specifically, managers and employees may resist this change and oppose the implementation of automatic systems capable of complementing human intervention in decision making. Thus, for modern organizations, the importance of simultaneously fostering the development of technical, managerial and personnel BDA related capabilities emerged (LaValle et al., 2011). Specifically, all the people who will have to deal with BDA in the organization, should be capable of at least partially understanding the complexity of the infrastructure, the main methodologies of analysis, the potential effects on existing processes and the potential outcomes of BDA (Côrte-Real et al., 2017). To do so, the culture of the whole organization must be changed to accept BDA capable systems and/or processes (Rialti et al., 2018). Anyway, if managers and employees will develop enough BDA capabilities to get accustomed to big data and BDA systems, it will be possible to observe that the whole organization could become characterized by the so-called “big data culture” and start harvesting big data benefits (Frisk and Bannister, 2017).
Notwithstanding all potential difficulties, once in place, big data systems tend to have positive outcomes (Akter et al., 2016). Managers may indeed make decisions according to the insights that they gather from BDA systems, thus, improving the accuracy of their decisions (Santoro et al., 2018). Big data can offer managers the possibility of knowing their consumers better than ever. With big data, it is possible to predict individual consumer’s behavior and propose tailored offerings in terms of prices (Erevelles et al., 2016). Big data can also dramatically affect organization’s internal operation efficiency and BDA may prove extremely useful for the control of performance of business processes (Del Giudice, 2016; Acharya et al., 2018). Indeed, BDA systems may allow managers to identify bottlenecks in the production processes, inefficiencies in machine usage and wasting of resources. BDA systems may also be linked to better workforce utilization and may permit managers to better monitor the performance of each employee. BDA systems may also positively impact an organization’s ability to pursue collaborations with partners. Particularly, BDA systems may improve knowledge flow and facilitate sharing between partners as they are frequently built around jointly developed hardware architectures (Vera-Baquero et al., 2016). BDA may play a role in fostering the organizational capability of identifying and seizing new opportunities. With the newly extracted information, BDA capable systems can improve organizational exploitation and exploration capabilities and, consequently, ambidexterity (Riali et al., 2018). In short, big data is progressively influencing organizations’ competitiveness.

As a consequence of the impact of big data on organizations, pertinent literature has stressed that these may be positively linked to better performance (Gunasekaran et al., 2018). In particular, it has emerged that big data and BDA systems and capabilities, both impact organizational performance metrics (i.e. workforce utilization, supply chain efficiency, production processes efficiency) and financial performance metrics, which may improve over the time (McAfee and Brynjolfsson, 2012). However, as previously assessed, all of the positive effects of big data derive from the organizational decision and the organization’s acceptance of big data. Organizational processes, such as resource allocation, orchestration, and exploitation, thus, play an important role in the organization’s ability to reap the benefits of big data (Teece, 2009). From this perspective, dynamic capabilities have been frequently involved in research exploring the importance of big data for organizations (Wamba et al., 2017).

2.2 BDA, dynamic capabilities and performance
The notion of dynamic capabilities was originally coined by Teece, Pisano and Shuen in 1997. According to their seminal manuscript, the essence of the dynamic capabilities concept lies in the organization’s “ability to integrate, build, and reconfigure internal and external competencies to address rapidly changing environments” (Teece et al., 1997, p. 516). The “dynamic capabilities” are related to organization-wide ability to adequately and timely adapt to the changing environment by reconfiguring internal or external processes and resources, with the existing competencies (Eisenhardt and Martin, 2000; Gaur et al., 2014).

While it may appear that dynamic capabilities definitions link organizations’ reactions to improvisation (i.e. it may seem that organizations simply respond to changes by spontaneous re-organization of resources using existing skills), actually dynamic capabilities derive from the existence of “identifiable and specific routines” (Eisenhardt and Martin, 2000, p. 1107). Some organizational routines and processes are capable of diffusing the best practices within an organization (Hwang and Gaur, 2009). In this vein, Eisenhardt and Martin (2000) observed how organizational routines or processes may be broken down into smaller routines, or small processes, which are the “bricks” for forming a completely new routine or process. Consequently, once an organization has implemented the original routine or process, formed by several bricks, these bricks may be reassembled to form a new routine or process.
necessary to survive and succeed in the mutated environment (Nuruzzaman et al., 2018). This phenomenon is linked to the assumed importance of expertise conservation within an organization, and to the fact that the same knowledge base may be used in more than one situation (Popli et al., 2017). Using this, it is possible to assess whether an organization has developed dynamic capabilities and it has become capable of adapting to change by exploiting existing resources, processes, knowledge and existing routines deriving from the continuous repetition of a similar action (Gaur et al., 2014). Thus, it emerges that the dynamic capabilities theory extends to both, resource-based view (RBV) and KBV (Côrte-Real et al., 2017). Dynamic capabilities, in fact, posit that the competitive advantage is not only driven from organization’s ability to reconfigure resources, but also from the ability to re-arrange them purposively (and timely) based on existing knowledge (Gaur et al., 2014; Popli et al., 2017).

In the big data era, dynamic capabilities have frequently been adopted by scholars as a theoretical perspective to unpack how big data or BDA systems and capabilities affect an organization (Wamba et al., 2017). To understand the effects of big data, scholars need to focus simultaneously on three perspectives, namely: data as resources; recurrent routines, processes and capabilities to analyze big data; and the management of knowledge emerging from these data (Ferraris et al., 2018). First, scholars should always consider that big data are an information resource characterized by a multiple usability potential, i.e. big data may be utilized more than once to get different information to solve diverse problems (Erevelles et al., 2016). Second, the analysis of big data requires routines, processes and capabilities to turn such data into meaningful insights (Côrte-Real et al., 2017), where previous expertise from analysts and managers can also be helpful in increasing the efficiency of big data analysis (Zeng and Khan, 2018). Third, as the analysis of such data sets may generate huge knowledge flows that scholars always have to consider for proper management of the knowledge emerging from big data to create value (Ferraris et al., 2018). Accordingly, it emerges that the singular use of RBV or KBV would not be sufficient to completely interpret how big data and BDA system and capabilities create value. For instance, when a scholar is using only RBV, he/she will only observe how big data can create value for the new information at managers’ disposition. By doing so, he/she will be neglecting the importance of routines in big data analysis. Similarly, the exclusive use of KBV will merely allow a scholar to observe how knowledge flows deriving from big data may influence decision-making processes, but it would not allow him/her to consider big data as an information-related resource that can be to solve more than one problem. Contrarily, the use of dynamic capabilities theory will allow a researcher to unpack the outcome of big data by considering simultaneously how existing routines to analyze data may allow multiple use of such data sets and to diffuse knowledge to all the people in the organization.

In the light of this, it is understandable why scholars have used dynamic capabilities to interpret the ways in which BDA systems and capabilities generate competitive advantages. Scholars have used dynamic capabilities to explain why the use of BDA systems is based on re-application of routines, which are fundamental for generating new information to overcome rivals (Braganza et al., 2017). Indeed, scholars have observed that the development of organization-wide BDA capabilities may trigger BDA systems users into learning new routines to analyze different kinds of data over time. Such routines may be used by users in different analytical processes that may be vital to run BDA systems during a change (Rialti et al., 2018). It emerges that BDA systems equate to constant knowledge generation and diffusion, through which they allow managers and analysts to identify good opportunities and to reject not-profitable ones. Second, dynamic capabilities have been used as a theoretical approach to observe how big data can affect marketing strategies (Khan and Vorley, 2017). This phenomenon has been deemed to be related to the potential of BDA systems to explore the behavioral patterns of consumers and, therefore, to foster the
creation of customized marketing strategies in a timely manner (Erevelles et al., 2016). Third, Wamba et al. (2017) have observed that the process oriented dynamic capabilities and BDA may influence both organizational and financial performances. In particular, they have observed that the possibility for a process to adapt to changing situations may be influenced by the diffusion of BDA systems within an organization. This is coherent to the fact that BDA information decision makers may predict what is going to change in the environment and accordingly modify processes. Similarly, Sivarajah et al. (2017) assessed BDA systems capacity to adapt to different kinds of data and to the evolving environments and deduced that this may generate competitive advantages.

According to pertinent literature, there is still a need to properly understand the structure and the organization of existing literature concerning big data, dynamic capabilities and performance (Braganza et al., 2017; Côrte-Real et al., 2017). Indeed, while several manuscripts have used the bibliometric method to explore other streams of big data related literature (Mishra et al., 2017), scant attention has been paid to this specific topic.

3. Methodology

Bibliometric methods have been widely used to provide comprehensive maps of the knowledge structure in a given streams of literature. However, as the authors are investigating an emerging field of research, to perform an accurate analysis of the literature, both bibliometric analysis and systematic literature review techniques are used (Caputo et al., 2018; Marzi et al., 2018). A bibliometric analysis was conducted first, followed by a systematic literature review of the bibliometric results. The bibliometric analysis is based specifically on the “visualization of similarities” (VOS) technique (Van Eck and Waltman, 2010). For the systematic literature review, the authors followed the procedure proposed by Tranfield et al. (2003). The entire process consisted of six steps.

The first step is the search of a wide research query on the Clarivate Analytics Web of Science Core Collection database, which offers the most valuable and high-impact collection of data and is recognized as the most updated and reliable database for bibliometric studies (Falagas et al., 2008). The process of selecting a research query began with a literature review of the cornerstone manuscripts about BDA capable systems for management, using dynamic capabilities as the main underlying theory to grasp all of the terms used to describe the phenomena that the authors wanted to analyze (i.e. Akter et al., 2016; Wamba et al., 2017). After several iterations to define a broad research query, the final query was:

\[ \text{TS} = \left( \text{“big data” OR “big data analytics”} \right) \text{AND} \left( \text{“dynamic capabilities” OR performance*} \right) \text{AND} \left( \text{organization* OR firm* OR business* OR enterprise*} \right) \]

The “TS” operator performed a full search of the selected terms in titles, abstracts, and keywords. The search was limited to “articles, books, book chapters, book reviews, early access articles, and editorial material,” as document type. A ten-year cross-section – 2007 to 2017 – was the considered as the timespan. A preliminary data set of 375 entries was generated by the query.

As previously assessed, research data were extracted only from Web of Science Core Collection database. In fact, Web of Science Core Collection, among the existing databases such as Scopus or EBSCO, has been frequently recognized as the database which includes most of the papers published in reputable journals over the time, including the majority of papers recently accepted by journals (Marzi et al., 2018). Additionally, if the same query is used and the same search parameters are set, previous research has pointed out that the use of Web of Science Core Collection usually provides less out of topic/aim papers that should be excluded from the analysis. Unlike Scopus and EBSCO, the Web of Science Core Collection does not include papers written in magazines or non-scientific journals (Caputo et al., 2018).
This phenomenon also proved to be true in this research. In fact, authors also manually checked Scopus and EBSCO databases and did not find any paper was not already included in the Web of Science Core Collection database results.

The second step was devoted to defining the inclusion criteria for the documents to be utilized in this study, and then to do the manual analysis and selection of each document. The authors decided to base the selection on three inclusion criteria, two of them related with the definition of big data and one related with dynamic capabilities. The first criteria were the most generally accepted definition of big data as “datasets whose size is beyond the ability of a typical database software tools to capture, store, manage and analyse” (Manyika et al., 2011, p. 1). The second criteria were the definition of big data as data sets simultaneously characterized by volume, velocity and variety, a.k.a. the original 3Vs of Big Data (McAfee and Brynjolfsson, 2012). Only the original 3Vs of big data were selected since the additional 4Vs (veracity, value, variability and visualization) were identified as characteristics of big data by scholars only in recent years (Wamba and Mishra, 2017). Considering the 3Vs was, therefore, an appropriate parameter as it was allowing to evaluate whether a research was effectively outlining big data using a widely accepted definition, without preventing older research manuscripts from being included from the data set. In the third criteria authors excluded manuscripts that did not consider dynamic capabilities as a research perspective and manuscripts not belonging to management-related literature. After applying these three inclusion criteria, the final data set consisted 170 entries.

The third step consisted of critically reading the 170 selected manuscripts by all four authors to obtain a working knowledge of how BDA are linked to dynamic capabilities and firm performances (Wamba et al., 2017). Subsequently, the fourth step consisted of the initial part of the bibliometric analysis. Specifically, the authors performed an analysis using activity indicators to gather data on the volume of research, allowing us to observe the quantitative evolution of the literature.

The fifth step involved the proper bibliometric analysis. Software tool VOSviewer 1.6.5 was used for the aggregation of the manuscripts, with bibliographic coupling as the aggregation mechanism (Van Eck and Waltman, 2010). Bibliographic coupling occurs when two works cite a common third work in their references; consequently, two documents are bibliographically coupled when they cite one or more documents in common. The output of VOSviewer is a map in which the items’ distance can be interpreted as an indication of the relatedness of the terms. The smaller the distance between the terms, the more strongly the terms are associated with each other. In addition, the cluster analysis highlights the knowledge base diversity in an aggregate way: if the manuscripts belong to the same cluster, it means that they are strongly linked together as a group based on their shared references. Thereby, a cluster represents a stream of research on a similarity basis. It is important to note that, on the map generated by VOSviewer, the manuscripts are presented in a convenient way to optimize their visualization; thus, the axes of the map do not have any meaning.

Finally, the sixth and last step involved systematic literature review process based on the results of VOS aggregation (Gaur and Kumar, 2018). Using the results of clustering found by VOSviewer, the authors analyzed the most influential manuscripts inside the displayed clusters to highlight their main areas of focus.

4. Bibliometric analysis’ results
In this section, the authors present the results of the aforementioned bibliometric analysis. The manuscripts’ distribution over the years is presented in Table I and Figure 1.

As shown, the majority of the selected manuscripts are five-years old or less. Only two manuscripts have been published in 2012 and no manuscripts were published prior to this. While research has explored the importance of big data and BDA for management since the
previous decade, only in the last five years, scholars started to explore this area using dynamic capabilities as a theoretical principle. According to the pattern revealed by the number of manuscripts published every year, the topic has yet to reach maturity. Indeed, the number of manuscripts on the topic are increasing every year.

In Figure 2, the results of the VOS analysis are presented. Due to space-constraints, only the most influential manuscripts are shown and only the surname of the first author is included in the figure.

From our analysis of the 170 manuscripts, four clusters emerged. As selected by the query, all of the manuscripts included in the clusters use dynamic capabilities as a theoretical principle. The content of the four clusters will be explained in the next section.

5. Systematic literature review
Coherently with previous research containing both a bibliometric analysis and a systematic literature review, the final part of this research comprises of a systematic literature review. Yet, as it was not possible to do a complete review of all the 170 papers, only the ten most cited manuscripts from each cluster were reviewed (Caputo et al., 2018).

5.1 Red cluster: big data, dynamic capabilities and supply chain management
This cluster aggregates manuscripts exploring the effects of big data on supply chain management, related dynamic capabilities and performance. Specifically, two groups of

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of papers</th>
<th>Variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>7</td>
<td>+250.00</td>
</tr>
<tr>
<td>2014</td>
<td>14</td>
<td>+100.00</td>
</tr>
<tr>
<td>2015</td>
<td>23</td>
<td>+64.29</td>
</tr>
<tr>
<td>2016</td>
<td>30</td>
<td>+30.43</td>
</tr>
<tr>
<td>2017</td>
<td>94</td>
<td>+213.33</td>
</tr>
<tr>
<td>Total papers</td>
<td>170</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ elaboration

Table I. Number of papers among the years

![Manuscripts' temporal distribution](source: Authors’ elaboration)

Figure 1. Manuscripts' temporal distribution

Big data and dynamic capabilities

2059
manuscripts emerged in this cluster. The first one deals with the implementation of BDA systems for supply chain management and the second is about the effect of big data and BDA on supply chain management performance.

In manuscripts about the implementation of BDA systems, Hazen et al. (2014), stressed that such systems’ managers may gain visibility into supply chain processes, costs and revenues or flows of materials. Specifically, the implementation of BDA systems may allow managers to obtain accurate predictions about the future needs of productive factors. Yet, such benefits will emerge only in the case when BDA systems can rely on high quality data; thus, BDA systems should be capable of depurating raw data and filtering out not-useful information generating noise around useful data (Kwon et al., 2014; Kache and Seuring, 2017). In fact, two manuscripts point out how BDA systems may cause positive effects only if they analyze high quality data. Actually, the usage of low-quality data (or data not cleaned of unworthy information) has been found to give erroneous predictions leading to wrong decisions. This notwithstanding, the risks related with low quality data are not related only with the quality of data analyzed by BDA systems. Kwon et al. (2014) also stressed that people may refuse to properly use BDA systems. It was observed that people may oppose a technology when they are not able to understand the potential benefits. Both Kwon et al. (2014) and Chaffin et al. (2017) have assessed that conducting training programs to develop organization-wide BDA capabilities may help. On this topic, Lavertu (2016) has mentioned the importance of external help (i.e. specialized consultants) for any kind of organization wishing to implement BDA systems to support supply chain management.

In case, the organization, as a whole, is willing to accept BDA systems for supply chain management, they will be capable of improving supply chain performances.

The second group of manuscripts, as stressed by Gunasekaran et al. (2018), states that the BDA systems may dramatically affect supply chain management performances. Specifically, it became apparent that the routinization of processes derived from BDA systems usage may occupy a fundamental role in ensuring the adaptability of these processes to different situations. Hence, BDA systems may increase organizational dynamism, agility and flexibility; particularly in the identification of problems and
opportunities related to supply chain management (Chen et al., 2015). The new information derived from BDA systems is indeed related with organizations’ capability to respond to changes that may disrupt supply chain (Dobrzykowski et al., 2015). According to Papadopoulos et al. (2017), BDA processes showing their ability to improving supply chain management have also been related to increased sustainability. This phenomenon is linked to the fact that increased efficiency in supply chains may lead to a reduced quantity of waste. Tan et al. (2015), has highlighted the potential of information from BDA systems for supply chain innovation. Specifically, they have highlighted how such systems may allow the exchange of information with supply-chain partners, thus, enacting the emergence of new innovative idea.

5.2 Green cluster: BDA method for knowledge extraction

The second cluster is formed by manuscripts about the importance of methods to extract knowledge from big data and exploit it. Lee (2017), for example, stressed that one of the biggest challenges of big data era is how to extract the needed information from big data and to turn it into exploitable new knowledge. Specifically, the aforementioned manuscript generically reviewed the most frequently used methods to solve big data related problems from the point of view of managers. In this vein, Zhou et al. (2014) focused on information technology methods that could make BDA systems work. As an example, it was emphasized that machine learning techniques are fundamental to analyze big data. Unsupervised statistical methods, allowing computers to automatically identify the most important insights, are in fact fundamental to process huge unstructured data sets such as big data. The importance for modern businesses to employ a specialized data analyst was also measured. Moving from these premises, Chen et al. (2014) and Li et al. (2016) explored how machine learning or artificial intelligence can be integrated into BDA systems. These manuscripts analyzed the characteristics of BDA systems, apart from the importance of machine learning and artificial intelligence, show how BDA systems should also rely on cloud storage and cloud computing as the size of big data has made traditional hardware obsolete.

Tirunillai and Tellis (2014), instead, have focused on techniques that may be used to generate knowledge for marketers. Specifically, they investigated the potential of Latent Dirichlet Allocation, which is a machine learning based topic classification methodology, to extract insights about consumers’ perceptions form user generated contents. In a similar fashion, Fuchs et al. (2014) and Yang et al. (2014) have stressed the simultaneous importance of web crawlers/scrapers (i.e. applications to scrape web pages or social media to collect data) and content analysis methods to predict demands for a service. Finally, Kwon and Sim (2013) have identified the potential of classifications algorithms to extract meaningful knowledge from big data. In any case, this manuscript explains the importance of re-thinking traditional classification algorithms to the new dimensions of big data.

Hence, for BDA systems to be able to extract knowledge from big data and increase performance, it was observed that the BDA systems may represent a fundamental tool to extract knowledge, know more about consumers and competitors (Chen et al., 2013; Al Nuaimi et al., 2015).

5.3 Blue cluster: big data, dynamic capabilities, decision making and performance

Due to the knowledge that big data can contain, scholars paid significant attention to the impact of big data on decision-making processes and subsequent organizational performances.

Fawcett and Waller (2014) and Tambe (2014) have recognized how BDA systems may generate predictions about future trends and that these predictions concerning sales, revenues and production requirements may be used by managers to formulate decisions about the future. Insights emerging from BDA can offer opportunities to managers to know
more about their consumers in real-time. Erevelles et al. (2016) have examined how BDA can enhance marketing related strategic decision-making processes. This phenomenon has been verified as the information about consumers allows a manager to react dynamically to evolving consumer preferences. Similarly, Opresnik and Taisch’s (2015) manuscripts showed how insights from BDA may influence the abilities of services providers to better tackle the needs of their consumers. Akter et al. (2016), Wamba et al. (2017) and Martin et al. (2017) pointed out how such insights may empower managers to take decisions that increase organizational efficiency in the supply and production processes. In fact, through predictive analytics it will be possible for managers to purchase just the minimum quantity of resources needed to cover the predicted request for products. Additionally, information from BDA system may also be fundamental to identify bottlenecks in production processes and reduce the wastage of resources.

Big data and BDA systems can, therefore, generate knowledge flows that are capable of changing the way managers think and act. Specifically, such flows may inform managers to take decisions about supply, production and sales. Managers may also be able to respond to sudden change as they may know almost exactly what is going to happen outside and, consequently, decide the best path for the organization to follow. In this perspective, Côrte-Real et al. (2017) evidenced that big data availability and the implementation of BDA systems represents a potential source of information-driven competitive advantage. Tallon et al. (2013) have highlighted that organizations may need to develop ad-hoc BDA system governance processes. In fact, such processes may improve the way managers access information to be used for decision-making. Similarly, it may allow managers to receive the right information for the right purpose. This is coherent to what was stressed by Constantiou and Kallinikos (2015), but, only when the output of BDA systems is aligned with managerial requests. Such systems can provide managers with the proper information to formulate adequate strategies.

5.4 Purple cluster: BDA, dynamic capabilities, business processes management and performance

The fourth cluster aggregates manuscript on big data, BDA systems and business processes management. From this perspective, Sivarajah et al. (2017) reviewed BDA methodologies and their effects on production processes management. They have shown that the possibility for managers to control any aspect of production processes is related to less wastage of resources. Indeed, BDA can provide managers a large amount of data that could allow them to make more informed choices on strategies (Wu et al., 2017).

Similarly, Vera-Baquero et al. (2016) have confirmed that BDA is related with superior performance by enabling manager to better monitor any internal processes. BDA managers may be fully aware of the performance of every process, identify problems or bottlenecks and sort-out the problem. Thus, by increasing the performance of each process, they may increase organizational performance. This topic is the talking point presented in two other manuscripts authored by Vera-Baquero et al. (2013) and Vera-Baquero et al. (2015). It has been observed that BDA systems’ analytical capabilities may not be matched by traditional business process management systems, as BDA systems are capable of providing managers a more detailed information in real-time.

Superior performance is the target for the majority of managers. As stressed by Gani et al. (2014) and Kowalczyk and Buxmann (2015), BDA may influence organizational processes to identify and exploit opportunities existing in the external environment. Specifically, BDA is capable of providing managers with the insights they need to formulate the strategies need to grab and exploit every emergent opportunity. This phenomenon is related with the accuracy of the real-time insights that may be extracted from big data. Nowadays, the perceptions and the ideas of consumers are no more a hidden treasure that
managers should look after, they are frequently freely available in the internet and, with the right tools, it may possible to analyze them. These insights may be diffused within the organization by the alignment between BDA systems and processes with knowledge management tools. As a consequence, strategic managerial decisions may now be supported by accurate information leading to increase in performance of organization by reducing costs or increasing revenues.

Nguyen and Cao (2015) have studied the ways in which BDA adoption may lead to stronger collaborations between partners belonging to the same production chain. The architecture of BDA systems may incentivise the sharing of data concerning production process efficiency between partners. In fact, BDA systems may also make communication processes occurring between partners more efficient.

Finally, Chae (2015) has proposed how the analysis of posts from social media can help organizations to better monitor demand trends. Thus, BDA systems may also affect demand forecasting processes.

5.5 Discussion of the systematic literature review and possible research gaps

The findings of the systematic literature review highlighted the importance of BDA for modern businesses. The possibility to apply advanced informatic and statistical analysis method is actually fundamental to make sense of big data and decodify their contents (Kache and Seuring, 2017).

In detail, results from the systematic literature review stressed out how BDA systems, tied up with organization-wide BDA capabilities to properly use them, matter to extract knowledge from data sets complex as big data (Dubey et al., 2018). Similarly, BDA systems may transmit information to interested players present within the firm. Hence, BDA systems may increase the quality and the speed of knowledge flows spanning the organization.

As a consequence, BDA systems and capabilities have proven fundamental to make managers obtain dramatically more information than before, particularly to what concerns any process occurring within the business and the supply chain (Mishra et al., 2018). BDA systems and capabilities may therefore allow managers to better decide about any future path the organization will have to follow.

In addition, why extant research used dynamic capabilities as the main theoretical approach emerged too. Indeed, modern organization to fully leverage BDA systems need to accept them, and need organization-wide capabilities allowing the organization to dynamically use existing systems for different scopes and with different kind of data (Ferraris et al., 2018).

To what concern theoretical findings of our research, several interesting topics emerged too. Specifically, first (as it is possible to see in Figure 1) it is possible to observe how clusters are extremely close each other. This may be related to the very close topics described in the considered manuscripts. Actually, the most of the considered manuscripts deal with the effect of BDA on manufacturing and supply-chain, thus they share many common references. Second, apart from few exceptions such as the manuscripts from Erevelles et al. (2016) and Martin et al. (2017), it is possible to assess that the most of research on big data and marketing, marketing management and marketing management scanty use dynamic capabilities as a theoretical approach. In fact, very few manuscripts dealing with big data and marketing are present in our clusters. Third, our analysis revealed how very scant attention is usually paid to two very important points. On the one hand, research has paid scant attention to factors fostering or hampering the adoption of BDA systems in modern organizations. On the other hand, very few attentions have been paid to the need for modern business to digitally transform to fully leverage BDA systems. Finally, there is a need to better explore additional potential effects of
BDA systems and capabilities such as increased innovativeness or increased absorption capabilities. As both these topics may be explored using dynamic capabilities, such an absence is significant.

Moving from this, the authors are also able to briefly suggest forward-thinking avenues for future research and possible gaps. Specifically, future researchers should try to explore, using dynamic capabilities:

1. the relationship between BDA systems and capabilities and dynamic marketing strategies;
2. organizational characteristics or organization-related factors that may hamper or prevent the adoption of BDA in modern organizations;
3. how organization may need to digitally transform to fully leverage BDA;
4. the effects of BDA on organizations structures; and
5. which are the additional effects of BDA analytics apart from better performances.

Apart from that, the authors observed that scholars should try to develop frameworks capable of reassuming the studied phenomenon. Additionally, as the majority of the research is theoretical or qualitative, quantitative research on the phenomenon is needed.

Moving from the aforementioned suggestions, we suggest managerial scholars wishing to contributes to this stream of literature to try also to collaborate with scholars operating in different disciplines, such as informatics, or practitioners. In fact, cross-collaboration may be helpful to provide different or unexpected insights on BDA systems and capabilities.

6. Discussions and conclusions

By identifying and reviewing the most influential manuscripts, the authors have systematized existing knowledge on big data and dynamic capabilities. This is the main theoretical contribution of this research. The authors have reconfirmed that four clusters exist in the research on dynamic capabilities (Wamba and Mishra, 2017).

From a practitioner-oriented standpoint, it is possible to conclude that managers should always monitor the alignment between big data capabilities and their expectations concerning BDA systems implementation (Akter et al., 2016). In fact, whether or not big data capabilities and organizational objectives are aligned, managers may find themselves without the insight they may need to develop new strategies.

In spite of these results, our manuscript is limited to a very narrow stream of research, and future research should therefore try to definitely systematise the position of the selected stream of research within a broader field.

References


Further reading


Corresponding author

Riccardo Rialti can be contacted at: riccardo.rialti@unifi.it

For instructions on how to order reprints of this article, please visit our website: [www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: permissions@emeraldinsight.com
Combining organizational change management and organizational ambidexterity using data transformation

Aditi Mitra
Department of Economics and Finance, Sunway University, Bandar Sunway, Malaysia
Sanjaya Singh Gaur
Department of Marketing, Sunway University Business School, Bandar Sunway, Malaysia, and
Elisa Giacosa
Department of Management, University of Turin, Turin, Italy

Abstract

Purpose – The purpose of this paper is to propose a practicable data-driven theory for the implementation and management of organizational change by combining the organization ambidexterity research and the organization change management research.

Design/methodology/approach – This study is based on the qualitative approach and uses a single case (in-depth investigation approach) study to come up with a data-driven theory, which is usable in the context of organizational change management and organizational ambidexterity (OA). Besides, in-depth interviews of change management practitioners, this study uses various sources of secondary information.

Findings – The study finds that owing to the reactive, ad hoc, and discontinuous nature of change often triggered by external factors or internal crisis within the organization, an organization need to continually engage with the existing data. The outcome must be driven toward preparing for the change through data engagement, implementation and reinforcement. The authors found that in order to be successful it is essential to have a strategy, set-up the right operating model, be clear on the scope of the change management work-stream and continuously monitor the progress through defined milestones and acceptance criteria. For companies targeting to achieve competitive differentiation through ambidexterity, a well-grounded change management program is the key for the success.

Originality/value – The study suggests that there is little work combining organizational change management and OA from a practitioner’s point of view. Accordingly, the authors propose a new data-driven organizational change management theory, which the authors term as the tripod theory for organizational change management. A practitioner’s perspective on the topic using a case study of an insurance company’s data transformation and a framework for structuring the change management program makes a meaningful contribution to the existing literature.

Keywords Data analytics, Organizational ambidexterity, Organizational change management, Data transformation

Paper type Research paper

1. Introduction

Change management is critical for the changing and dynamic business environment of today. It is considered as a critical organizational capability. Organizational change management (OCM) involves continuous assessment and renewal of an organization’s direction, structure and capabilities, in response to the changing demands of different stakeholders (Moran and Brightman, 2000). There is widespread recognition among leaders in most industries that the role of digital technology is rapidly shifting, from being a driver of marginal efficiency to an enabler of fundamental innovation and disruption (World Economic Forum, 2016). However, in order for companies to transition themselves to
becoming a digital enterprise there is a need for investments in digital technologies (Wamba et al., 2017). As part of digital transformation, organizations go through changes in business models, operating models, human talent and skill requirements. It is critical to get a close fit between digital transformation strategies, IT strategies and other organizational and functional strategies (Henderson and Venkatraman, 1993).

In similar light, it is critical to recognize the importance of data to initiate the digital transformation process (Fosso Wamba et al., 2015). Past studies (Gehrke, 2012; Yaqoob et al., 2016) have indicated the positive impact of data analytics on the firm’s resources and business capabilities (Acharya et al., 2018; Akter et al., 2016; Davenport et al., 2012). Organizations start their journey for digital transformation by first defining their enterprise data strategy. The objective of this is to address key challenges and to support the strategic objectives of the organization (Wamba et al., 2017). In the digital age, informed, data-base decision making is one of the key competencies that determines organizational success and failure (Gaur, 2006). Typical components of companies’ enterprise data strategy include building a single and central source of truth for data with confidence in the data and information we use and present; operational and governance models that provide clarity on the roles, responsibilities and services; fostering a data-driven mind-set with focus on using data to identify where problems or opportunities are to allow to make better decisions; and ability to have an integrated view of data and information that provides insights at an enterprise level. All these components need a dedicated focus on managing the change and ensuring that the benefits of strategic transformation are sustainable. There is a widespread recognition among both academicians and practitioners that any digital transformation starts with building the capacity to capture data about customers as well as partners and the ability to analyze the data to make informed decisions.

A new research paradigm called organizational ambidexterity (OA) has evolved wherein companies go through a dual transformation by exploiting current competencies and enabling differentiated capabilities through strategic transformation (Gaur et al., 2014). Preponderance of evidence shows a clear pattern: ambidexterity has been shown to be positively associated with the sales growth (Auh and Menguc, 2005), subjective ratings of performance (Bierly and Daly, 2007), innovation (Adler et al., 1999) and market valuation as measured by Tobin’s Q (Goosen et al., 2012). Organizations face an inherent confusion: whether ambidexterity is to be achieved sequentially or simultaneously. In today’s dynamic market with an ever decreasing innovation cycle, companies need to pursue the two activities together to remain competitive (Singh and Gaur, 2013). AlShaima et al. (2016) concluded that there is a linkage between enablers of knowledge and its sharing dimension. This linkage has a positive impact on the firms’ innovative capacity and processes. The degree of change management in handling such a dynamic ambidexterity is significant. This study combines the organization ambidexterity research and the organization change management research areas in the context of an insurance company undergoing data transformation to provide practical recommendations for success in the market place. The study, through in-depth interviews of change management practitioners and case studies of organizations identifies the key success factors for the management of such change. The study also provides a practical framework for a change management work-stream in a data transformation program (DTP) demonstrating its application in the context of an insurance company.

Considering that 70 percent of large change management programs fail and given that the rate of change has never been greater (Balogun and Hope-Hailey, 2004), there is a need for stronger emphasis on the topic, especially in the context of data transformation. A systematic review of the previous literature (Gaur and Kumar, 2018) shows there is a general lack of empirical research on change management within organizations, and an
arguably fundamental lack of a valid framework for organizational change management (Lines, 2005). Our review suggests that there is little work combining organizational change management and OA from a practitioner’s point of view. Therefore, a practitioner’s perspective on the topic using a case study of an insurance company’s data transformation and a framework for structuring the change management program makes a meaningful contribution to the existing literature.

The rest of this paper is organized as follows. Section 2 provides the literature review on change management and data transformation as well as the existing models. Section 3 proposes the methodology of this study. Section 4 presents a conceptual model for an enterprise data strategy and the supporting change management program for an insurance company. In Section 5, we apply the model for an insurance company and provide learning through sample outputs. Section 6 is for discussions and implications.

2. Literature review

2.1 Organizational change management

We conducted a brief but focused literature review following the guidelines provided by Gaur and Kumar (2018) for reviewing past studies on organizational change management (OCM) which mostly focus on the definition and application of existing change management models. One of the most popular and effective change management models have been Lewin's Force Field Analysis, also known as the three-step model of change management (Bumes, 2004; Armstrong, 2006). The three steps include unfreezing, changing and refreezing the processes and systems in a firm that is working to adopt a change. During the first stage of unfreezing, organizations attempt to alter the existing stable equilibrium which maintains present behaviors and attitudes (Armstrong, 2006). In the second step, there is cognitive restructuring such that the actors acquire information and evidence to support that the change is desirable and possible (Katz and Kahn, 1978; Schreyögg and Noss, 2000). In the third step of refreezing, organizations again achieve a new equilibrium, as all changes in the transformation stage are made permanent (Cummings and Worley, 2001).

Another important model on change management is Beckhard’s (1969) Change Plan, which incorporated the following processes. First, organizations need goal setting and defining the future organizational situations desired after the change. Second, one needs to assess the current conditions in relation to the desired goals. Third, organizations need to define the transition state activities and commitments required to meet the future state. Finally, organizational actors need to develop strategies for managing the transition based on the analysis of different aspects that are expected to influence the beginning of change. Building on this, Thurley (1979) introduced a change model that includes five major strategies to manage change: directive, bargained, hearts and minds, analytical and action based.

Kotter (1995, 1998) developed a model which is appropriate at the strategic level to change an organization’s vision and subsequently transform the organization. Kotter’s model included eight steps: first, creation of a sense of urgency for the change; second, forming a powerful coalition of managers to work with the most resistant people; third, creation of a plan comprising of vision and strategies to accelerate the change; fourth, communicating the vision to help people know that change is near which makes them less likely to resist; fifth, encouraging and inspiring people to adopt change; sixth, planning for and creating short-term wins; seventh, gathering feedback and consolidate improvements; and eighth, institutionalizing the changes.

Another model, known as the Continuous Change Process Model (Tichy and Ulrich, 1984) looks at planned change from the perspective of the top management and considers that change is a continuous process. The model focuses on change agents and the need for an organization to seek their assistance and make them responsible for managing
the change. The model also focuses on communication of the changes to all stakeholders, including employees, customers and suppliers (Tichy and Ulrich, 1984). Finally, action research model of change is a combination of changing not only attitudes and behavior, but also testing the change method being utilized (Collier, 1945; Lewin, 1945, 1951; Argyris, 1968, 1970; French, 1969; Schein, 1990; McShane and Von Glinow, 2005). It refers to change process based on the systematic collection of data and then selection of a change action based on what the analyzed data indicates.

The above review suggests that in much of the literature, change has been conceptualized in two fundamental ways. First, scholars see change as a rational strategic process where the organization chooses a new course of action and adapts to change. The second approach views change as evolutionary process, where organizations typically resist the change happening around them (Flood and Fennell, 1995). In real life situations, organizations often either adapt through strategic processes, or they fail to see the need for change and are replaced. Wiggins (2009) cited flawed maps of change, complex problems, superficial solutions, misunderstanding, resistance and misuse of knowledge about change management process as the main challenges in the change management process. Anyieni et al. (2013) further argued that change management involves planning, initiating, realizing, controlling and stabilizing the change processes at both corporate and personal levels. Since change often affects people, both inside and outside of the organization, many managers find it difficult to adopt changes Carr (2003).

Extant global theories and approaches for change management are often contradictory, and lacking practical evidence about the drivers and outcome of OCM. While there are studies that examine organizational change with reference to specific strategic choices that firms make, there is a lack of empirical research on change management itself within organizations and specific to sectors. For example, Singh et al. (2017) examined how firms engage in expansion projects in home markets when faced with pro-market reforms in emerging economies such as India. In a similar vein, scholars have examined international expansion as a strategic response to various internal and external pressures (Gaur et al., 2018; Gaur and Delios, 2015; Kumar et al., 2012; Popli et al., 2016, 2017). However, as argued before, these studies do not focus and emphasize on change management process within an organization. There is a clear lack of a framework that can guide a successful implementation and management of change within an organization. What currently available is a wide range of contradictory and confusing theories and approaches, which are mostly lacking practical evidence and often based on unchallenged hypotheses regarding the nature of contemporary organizational change management (Todnem, 2005).

Our review of the extant literature indicates only four studies focusing on OCM in the insurance sector. There is also an arguably fundamental lack of a valid framework for OCM in these papers. We address this oversight in this paper. Using case studies, we propose a framework for change management and identify the factors that lead to success of such efforts in organizations.

2.2 Organizational ambidexterity strategy

Tushman and O’Reilly (1996) were the first to present a theory of OA. According to them, the dilemma confronting managers and organizations is clear; in the short run they constantly increase the fit or alignment of strategy, structure and culture which is the world of evolutionary change but is not enough for sustained success. In the long term to be successful, managers need to embrace revolutionary innovation which may destroy the very alignment which has made the organizations successful in past. The number of studies in leading management journals that explicitly refers to the ambidexterity concept increased from less than 10 in 2004 to more than 80 today. This increasing attention has contributed to the refinement and extension of the ambidexterity concept.
Previous empirical research has investigated the effect of the exploration–exploitation dichotomy on performance from various perspectives, the implication being that both strategic acts may lead to different innovation performance outcomes (He and Wong, 2004; Katila and Ahuja, 2002; Lavie et al., 2010; Singh and Delios, 2017). For example, many studies reported the positive performance effects of the balance between exploration and exploitation (He and Wong, 2004; Jansen et al., 2006; Lin et al., 2007). Studies also exist indicating a negative effect (e.g. Lavie et al., 2011). Clearly, much clarity is needed on how ambidexterity affects organizational level outcomes.

A study of 80 studies on OA suggests several gaps in the existing research. First, there is no clear conceptualization of the OA construct. Second, there are very few studies that provide a practitioner’s view on the OA construct. Few studies that examine this issue are mostly conceptual, with no application of the concepts in a real organization's context. Third, OA in prior studies is to a large extent influenced by specific methodological choices adopted by the researcher. Furthermore, studies provide mixed empirical evidence because studies on OA have been conducted using different measurements and research designs. Also none of the previous studies provide directions for managing change while achieving OA. Clearly, there is a need to combine conceptual and practical work to present a perspective for achieving sustainable performance through ambidexterity and managing the change.

3. Methodology
This study uses the qualitative approach to achieve a deeper understanding of the association between organizational change management and OA as well as the transformation program in a firm. This study uses a single case (in-depth investigation approach) to come up with a conceptual model that is usable in the context of organizational change management and OA. Yin (1984) defined case study as a research method which requires in-depth investigation about the topic which may be new and more suited in the real-business context. The proponents of the case study approach have been advocating to gain deeper understanding about a new or complex topic (Bonomo, 1985; Buhais and Antonella, 2000; Halinen and Tornroos, 2005) or to strengthen the existing sources of information (Parkhe, 1993; Yin, 1994, 2003). Yin (1984) also concluded that the existing analysis of the documents is the starting point for the case study approach. Accordingly, our study uses the case study research approach through integrated evidence collection (i.e. using past documents of a firm), interviews and other secondary sources (such as academic journals, magazines and company websites). This study aims to provide insights on the topic by answering the questions such as what is the basis for a transformation program, how can organizational change management be achieved and how firms can overcome OA at the end of the transformation program? This is of utmost importance in today’s business environment where the dynamics of business functions and customer expectations are always changing.

4. Conceptual model
One of the more enduring ideas in organization science is that an organization’s long-term success depends on its ability to exploit its current capabilities while simultaneously exploring fundamentally new capabilities (Levinthal and March, 1993; March, 1991; Hwang and Gaur, 2009). A well-grounded data strategy for any organization is to build the organizational capabilities to exploit its core business, while at the same time exploring new opportunities through a strategic transformation.

In the context of data transformation for any organization, there are significant opportunities to leverage data to create an ambidextrous organization. Opportunities are present across the value chain to strengthen the core and to create and explore new opportunities to refine the competition. Overall, we argue that data can assist in knowledge
co-creation, which can, in turn, adequately lead to evidence-based, effective and efficient decision making for better business returns (Acharya et al., 2018).

In the context of this study, we have investigated an insurance organization using the legacy workflow systems. The organization is termed as the “legacy insurer” to protect its identity. Kenealy (2012) concluded that insurance companies around the globe use legacy transformation systems. Hence we use this reiteration to provide rationale for using the term legacy insurer for the purpose of our study. Unlike enterprises that are born digital, traditional companies, such as legacy insurer need to build a platform particularly designed for the digital enterprise on a legacy foundation. Based on Figure 1, one can say that organizations have historically pursued innovation in their core business. However, future leader needs to consider the following:

- today’s engine (exploitation) – looking for sustaining innovations as efficiently as possible in current business situation; and
- tomorrow’s engine (exploration) – that reflects new customer needs, new competitors and new economics.

A clearly defined data strategy will help companies develop their customer facing capabilities while decoupling legacy systems. In general, there are four major pathways leading insurance companies have taken to transform themselves into a digital organization (see Figure 2). These pathways orient a legacy insurer and enable them define their competitive advantage. None of these pathways are inherently better than others, and they are not mutually exclusive. Nevertheless, it is imperative for a legacy insurer to assess the internal capabilities and establish a clear strategic direction in parallel. As organizations transform, aligning data strategy to the strategic direction of the organization is becoming a critical activity. In many cases, ability to discover insights within the data, orchestrate resources around the discovery and create new value proposition are the key differentiators that help achieve the state of ambidextrous organization. Past studies also indicate that data-driven decision making leads to the output and productivity that is 5–6 percent higher than what would be expected given their other investments and information technology usage (Brynjolfsson et al., 2011).

As is evident from Figure 2, the digital pathway of “Advanced Analyzer” which primarily focuses on building data capabilities in any organization helps to deliver high “customer knowledge” and deliver high “data-driven quotient.” In the current business environment, companies have higher ability to differentiate when they understand their customers/partner better and can deliver propositions aligned to their specific needs.
For any company aspiring toward being “Advanced Analyzer,” the fundamental scope of their transformation is centered around being “Data Aggregator” and “Data Analyzer” across all the transformation initiatives being executed.

Seizing the data-driven opportunity provides a unique opportunity for legacy insurers not only to strengthen and exploit their competencies but also to provide differentiation by exploring new opportunities. In Figure 3, we present examples of how legacy insurance and technology companies are exploring opportunities to achieve the positioning in an ambidextrous way.

Transforming any legacy insurer into a data-driven organization requires significant technological, cultural, people and process change throughout the organization. The operational processes need to be completely changed such that the organization becomes less dependent on the past experience of the leaders. As mentioned earlier, the scope of change is more significant for companies targeting to achieve ambidexterity and an appropriate change management program therefore becomes an imperative.

The overall approach toward change management is carried out in three phases: preparation, implementation and reinforcement. The execution of a change management program is carried out both at an enterprise level and at individual level. Acceptance criteria need to be defined for the change management work-streams both at the project and at the program levels. The preparation phase comprises of the analysis related activities and the development of the transition plans which are subsequently implemented in the next phase. In the reinforcement phase, the realization of the program and change management benefits is measured. Mechanism is established in this phase to gather feedback and take appropriate actions to continuously improve the performance of the organization. This is further explained with the help of Figures 4 and 5.

5. Role of data transformation in legacy insurance company to support organizational change management
As mentioned earlier, this paper investigates the data transformation of a mid-sized legacy insurance company and provides an application of organizational change management
framework for a DTP aimed at achieving the ambidextrous state. In-depth interviews that have been conducted during the course of the study to understand various facets of changes across all the dimensions. Based on this, a practical approach to implement the change management work-stream has been put forward. The outputs provided in the study are not specific to any legacy insurer but are based on the reflection of the common findings which are applicable across any legacy insurer.

5.1 Analysis of the existing legacy insurer

The model presented in Figure 6 collates observations that provide a high level understanding of the current state of the legacy insurer.

The model to understand the existing legacy insurer and the matrix indicating the need for change among legacy insurer have been provided in Figures 6 and 7, respectively. Based on the case study method and analyses of the exiting information, the matrix indicating the need for change among the legacy insurer has been developed. The matrix indicating the proponents of the need for change among legacy insurer also links it to the significance of a DTP. In the context of this study, the DTP has been initiated and undertaken by the particular insurance company to successfully spearhead OCM and overcome OA. DTP is part of the overall company’s digital transformation mandate and needs to be aligned to deliver the digital program’s target benefits. Broadly the data transformation would need to address the challenges associated with the current state, ensure compliance and risk management adherence, deliver operational effectiveness and develop competitive advantage through achieving ambidexterity (see Figure 7).

To achieve the data transformation benefits and address the challenges associated with current (existing) state, the legacy insurer needs to embark on a DTP which comprises of four...
major building blocks. The work-streams within the DTP need to have an alignment with the overall OCM program the company has embarked upon. For this the company needs to:

- build data and analytics organization: this block comprises of all activities related to establishing the operating model and team, as well as implementation of governance practices;
• uplift technology: the block comprises of building the foundational technology related to initiatives such as enterprise data platform, etc;
• deliver reporting and analytics capabilities: in an agile manner, this block delivers the reporting and analytics capabilities to the enterprise from the enterprise data platform; and
• manage change: in this transversal work-stream, effective and smooth transition is ensured through planning of change management.

5.2 Organizational change management – Phase 1: preparation
Figure 8 depicts operating model for the organizational change management. In the following section the organizational change management (OCM) approach and the key deliverables are presented for the DTP. The change management program becomes more significant as it needs to align to the overall digital transformation change management work-stream and needs to be dynamic enough to achieve ambidexterity.

5.2.1 Establishing the change management operating model. The OCM is initiated by setting-up the overall change management strategy and putting in place the transversal and project specific change management organization. The transversal change management is across initiatives and reports to the program execution head and the program sponsor. The project level change management consists of key resources within the initiative and plays an important role in
the operationalization of the strategy. Many organizations deploy external consultants to operationalize change management programs. Such program consultants usually provide only a strategy/framework but the employees in an organization have to take the responsibility for rolling out the programs (Arvin, 2014) and often they do not get proper guidance to do so.

5.2.2 Assessment of stakeholders’ needs. The OCM process starts with identifying the stakeholders impacted by the data transformation. Involving employees as part of the implementation process facilitates the transition toward change management as employees become familiar with the change (Gupta et al., 2018). The study by Al Mansouri et al. (2018) supports this preposition where they argued that the architecture of knowledge management is dependent on the internal culture and leadership style of a company, the prevailing structure of the firm, existing leadership of the firm and the overall companies’ citizenship behavior. Thus, concluding that OCM has to consider both firm level and individual level input variables of the firm. Analysis is done on the existing state by identifying where does the data capabilities lie within the organization. Considering the data transformation is transversal in nature, the impact is organization wide but the degree of impact would vary based on different stakeholders’ data needs. In addition to the stakeholder identification, the restraining factors are also identified and initial hypothesis is developed for resistance management approach. One approach to understanding these risks is to capture each stakeholders’ perception of the change through semi-structured interviews allowing stakeholders to talk about the benefits, risks, opportunities or concerns as they might perceive them (Vidgen, 1997; Kambil and Heck, 1998). The hypothesis is subsequently refined within the “organization development and culture” work-stream as organizational culture plays a significant role in understanding the resisting forces. Indicated in the Figure 9 is the way a typical legacy insurer is assessed currently. The distribution of the activity types in the assessment of any data organization is also indicated in this figure.

A stakeholder impact analysis is performed to understand the degree of impact of the data transformation change on the functions and the degree of influence the function has on the success of the organization. The assessment is performed through in-depth interviews, surveys, focus groups, etc. Based on the evaluation, a typical stakeholder impact map is drafted and is indicated as shown in Figure 10. For organizations targeting to achieve ambidexterity, the assessment is significant as it provides the initial hypothesis for the impactful resistances to change and aids to build a resistance management approach. Also while organizations prepare for achieving ambidexterity, there are new functions, roles and responsibilities which need to be created but are not in place yet. Hence, the stakeholder assessment needs to incorporate the future operating model too.
5.2.3 Change impact assessment. The change impact analysis is performed to identify scope and scale of change management needs for the organization to become data driven. This step serves as an input for the formulation of the change management plan. The assessment of the impact is performed on four dimensions: people, process, technology and policy with each change impact classified into exploration or exploitation type. During the assessment, the key challenge is to identify the impact of change due to the exploration of new opportunities by the organization. For this step to work, it is recommended to work with experienced change management practitioners who in addition to having an understanding of the change management principles also have a strong understanding of the transformation context and the changes it would entail. A typical change impact assessment worksheet is provided in Figure 11. Subsequent to the listing down of the change impact, change agents are identified within the organization and made accountable to deliver the target state of each change impact with clearly identified deliverables and timelines. Dependency between the change impacts is accounted and the timelines are drafted accordingly.

5.2.4 Capability gap assessment. A key part of the process is to determine if the energy required for the change can be mobilized by testing for organizational readiness (Benjamin and Levinson, 1993). Based on the inputs of the stakeholder mapping and the change impact, a detailed capability gap assessment is performed considering both the axes of
change: exploitation and exploration. Considering the significance of this step, it is recommended that a detailed individual level assessment is carried out. The target capabilities are listed down to achieve the proposed future state of the data-driven organization. An individual level assessment is also carried out to assess the significance of the required capabilities and the level of existing competencies available for these capabilities. This assessment provides recommendations for the development strategy and to ascertain that the required capabilities can be built internally through up-skilling and re-skilling (or by acquiring from external sources). A key decision during this step for organizations is to formalize the capabilities which are core to transformation. It is always recommended to build required competencies in-house through investing in adequate resources through need-based training to stay competitive and to improve job performance (Al Mehrzi and Singh, 2016). The importance of job performance need not be stressed (Bozionelos and Singh, 2017). Pradhan et al. (2016) concluded that in the current changing business environment, companies need to focus on securing adaptive performance from their employees. The process of adaptive performance can be initiated when there is a synergy between emotional intelligence and the firm culture. Likewise, for non-core capabilities the skillset is recommended to be sourced on a need basis. Listed in Figure 12 are few of the capabilities essential for any legacy insurer to achieve the data driven target state with the capabilities being further aligned to the transformation axes.
5.3 Organizational change management – Phase 2: implementation

Second phase of OCM deals with implementation of the plans developed in Phase 1. Figure 13 depicts the OCM implementation phase. Cultural transformation, organizational development as well as communication, knowledge management and training and development are the key aspects of this phase of OCM. These are discussed in detail in the following section.

5.3.1 Cultural transformation. Senior executives of the most organizations widely believe that cultural change builds the foundation upon which all the other strategic initiatives rest. The success of an organization depends on the competencies of its leaders and the organizational culture those leaders create (Al Matrooshi et al., 2016). The development of more efficient and effective processes and alignment of organizational culture to support these new processes are critical for successful change to occur. If companies succeed in changing their leaders' and managers' mind-set, the rest should follow relatively smoothly. Being data driven requires a significant cultural shift from making decisions on intuition and experience to making decisions on hard facts. Therefore, the need for a cultural reboot to ensure success in any data transformation for an organization is paramount. This process of cultural transformation starts with an understanding of the key challenges associated with the mind-set of leaders and managers in an organization. For any legacy insurer, the typical current and target states for organizational culture are shown in Figure 14.

For building a data-driven cultural mind-set, first people need to “know their data.” For this, an enterprise business dictionary/catalog is developed and mapped with the baseline data so that a consistent definition of all the data elements across the enterprise is understood. Once the enterprise data is baselined, then the process of brainstorming can be started to understand how employees within organization can become data driven.

5.3.2 Organization development. In this step, the target operating model for the data and analytics organization is established. The roles and responsibilities are also defined in this step. Further human resources are identified either within the organization to assume the roles or identified for sourcing externally. The capability gap assessment drives the skillset requirement and is attributed to each of the role within the target operating model.
Many organizations separate their new, exploratory units from their traditional, exploitative ones. This allows them to have different processes, structures and cultures. At the same time, they maintain tight links across units particularly at the senior executive level (Tushman, and O’Reilly, 1996). Such “ambidextrous organizations” allow executives to pioneer radical or disruptive innovations while also pursuing incremental gains.

Different target operating model options exist for organizations and are listed in Figure 15. For a typical organization, each of the models listed below have their pros and cons. Therefore, the final selection is driven by the future vision as well as the current state, maturity and size of the organization.

Figure 16 depicts the recommended target operating model for a legacy insurer. The final adopted structure is hybrid in nature to ensure close alignment with all the business functions.

<table>
<thead>
<tr>
<th>Option</th>
<th>Team structure</th>
<th>Characteristics</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
</table>
| 1      | Decentralized  | ▶ Maintain status quo  
▶ Business unit control reporting  
▶ Ability to understand specific business unit needs  
▶ No shared service | ▶ Responsive  
▶ Business unit specific knowledge  
▶ Ownership within business units | ▶ Duplication of effort  
▶ Limited corporate view of customer, processes, outcomes  
▶ No central ownership |
| 2      | Centralized    | ▶ Reporting and Analytics center of excellence  
▶ Business and technical resources centrally pooled  
▶ Provide services to business areas as requested | ▶ Efficiency and clear demand pipeline view  
▶ Corporate picture  
▶ Central Ownership | ▶ Not tightly integrated with end users  
▶ Slowest response time  
▶ Perceived loss of control by senior execs |
| 3      | Hybrid         | ▶ Reporting and analytics expertise reside within centralized area  
▶ Business resources reside within business area  
▶ Business resources move the centralized area for defined projects  
▶ Establishment of self-service reporting | ▶ Better communication and information sharing  
▶ Expertise shared across organization  
▶ Development of a standardized view of data definitions  
▶ Centralized but superior to centralized | ▶ Perceived loss of control by senior executives  
▶ Requires additional executive sponsorship and activity  
▶ Response speed is not as fast as decentralized but superior to centralized |
5.3.3 Communication, knowledge management and training and development.
Organizational communication scholars have recently started to pursue new directions for organizational communication. These directions are very different from the instrumental and linear-like use of communication for achieving change goals (Frahm and Brown, 2007). Past experiences in transformation programs indicate that a change communication is considered effective only when various stakeholders can relate to the changes to their daily working and have a mechanism to feed up communication which traditionally came from above. Effective organizations conduct periodic workshops targeted at helping all employees relate to the changes and continuously gathering feedback so as to make everyone feel included. Considering data transformation requires an enterprise wide deployment, an effective communication strategy for any DTP is a key success factor. Effective implementation of knowledge management initiatives enables an organization to respond efficiently toward market changes and reduces the amount of redundant information available to the organization. For such organizations, knowledge becomes one of the most critical driving (Arvin, 2014) force for business success (Kuan, 2005). Eventually, in such organizations, change management teams work closely with the human resource teams to deliver the knowledge management initiatives and complement these initiatives with well-structured training and development programs.

5.4 Organizational change management – Phase 3: reinforcement
With the implementation of the change management initiatives, the work-stream does not end. There is a need for defining the acceptance criteria (during the preparation phase) for the initiative level as well as the program level. After implementation, a mechanism needs to be established to monitor the change management performance. Similarly, for each of the initiatives as well as for the program, expected target benefits need to be articulated.
It is within the purview of the change management work-stream to establish a mechanism and continuously monitor the progress. For this to happen, a continuous feedback mechanism needs to be established and steps need to be taken for continuous improvement. It is applicable for both initiatives as well as program. The role of the change agents is essential even in the reinforcement phase wherein they continue to assess the post implementation impact and continue to feedback to the change management organization. Many organizational scholars (Kanter; 1983; Schein, 1987; Kotter, 1995) often have argued for the critical role of change agents. The strategic nature of transformational leadership role has been recognized as a critical change agent role because if its ability to create inspiring visions (Burns, 1978).

6. Discussions
In this section, we discuss the findings and the implications of our research. We also present avenues for future research and limitations of our study. This study aims to provide a guidance on how OCM may be adopted even though it may be unpredictable. The findings of this study are supported by the past literature (Burnes, 2004; De Wit and Meyer, 2005; Luecke, 2003; Nelson, 2003; Mukherjee et al., 2013) where authors have concluded that owing to the reactive, ad hoc and discontinuous nature of change often triggered by external factors or internal firm crisis, a firm need to continually engage with the existing data. The outcome must be driven toward preparing for the change through data engagement, implementation and reinforcement. We term this proposed theory as the tripod theory for organizational change management. Additionally, the paper concludes that poor success rate of OCM is due to the fundamental lack of a valid framework for the implementation and management of the organizational change. In a similar vein, Burns (2004) also concluded that it is due to conflicting frameworks proposed for the adoption of change management. Thus, to overcome this limitation, this study proposed the adoption of the tripod theory for data-driven OCM. Use of this theory can help firms overcome the resistance to change management programs. This tripod theory of OCM is proposed based on the inputs from change management professionals besides the literature. Therefore, it is a practical change management approach which provides users with insights across the overall lifecycle of preparation, implementation and reinforcement. This study also recommends that a DTP of a firm should be separated from the project management of a firm so as to utilize the synergistic benefits created by the change management work-stream (i.e. the project management team acts only as a support to the DTP). However, this synergy can be hindered due to budgetary constraints, lack of communication between these two teams, and unwillingness to work toward change management work-stream. These potential problems in transition may have a significant impact on the OCM process. Therefore, the leadership needs to ensure that there is sufficient knowledge sharing touch points among various stakeholders and transparency about the transformation program is maintained during the whole process.

6.1 Managerial implications
The key focus of our study was to provide guidance to managers and transformation consultants who are engaged in implementing changes in various organizations. While the context of our case study is that of an insurance company, findings can be easily implemented in the organizations in financial services sector as well as in the organizations with a traditional model of customer interface. Any transformation program, whether data based or otherwise, at the end is a change or transition program. Therefore, it is crucial for the success of the program that managers invest time and resources in establishing a strong operating model. In organizations, the change management work-stream is either neglected or is staffed with personnel who tend to focus on change management merely as a
communication and training and development exercise. In order to be successful, it is essential to have a strategy, set-up the right operating model, be clear on the scope of the change management work-stream and continuously monitor the progress through defined milestones and acceptance criteria. For organizations targeting to achieve competitive differentiation through ambidexterity, a well-grounded change management program is the key success factor. The focus of managers in such a case would be to continuously exploit the core capabilities and explore differentiating capabilities to stay ahead of the competition. They should build a change management program which ensures long-term sustainability of the program benefits.

6.2 Theoretical implications
As mentioned in the preceding section, this paper proposes a new data-driven theory which we term as the tripod theory for organizational change management. Application of this theory also enables organizations to overcome ambidexterity as a result of the transformation program. Since there is lack of a practical data-driven theory to guide OCM and manage OA, development of this theory is a significant contribution to the OCM literature and has potential to generate significant interest among researchers as well as practitioners.

6.3 Limitations and potential for future research
Since the proposed framework is based on a single case study in the context of an insurance providing service organization; we hope that it can be applied in the context of service organizations in general and particularly in the context of financial services. However, we admit that use of a single case study limits the generalizability of our findings. Therefore, empirical examination is essential to evaluate the potential applications of the proposed Tripod theory in various contexts including other service contexts. Further studies should emphasize on validating the proposed theory in terms of relevance, practicality and adequacy. Testing of the proposed theory across various organizational contexts and other types of transformation programs would provide beneficial information to professionals for employing the proposed framework. Future research should focus on both empirical and conceptual investigation for each of the phases: preparation, implementation and reinforcement as well as for the various aspects of each of the three phases. More such studies should enable an identification of critical success factors for the data-driven transformation based management of change. Furthermore, in order to construct a valid framework for change management, it is arguably necessary to enable measurement of the success rate of change initiatives. Further studies focusing on designing of the methods of measurements should, therefore, be planned.

References


Further reading


Corresponding author
Sanjaya Singh Gaur can be contacted at: sanjayag@sunway.edu.my

2091

Combining organizational change management
Big data analytics capability in supply chain agility
The moderating effect of organizational flexibility

Rameshwar Dubey
Montpellier Business School, Montpellier, France
Angappa Gunasekaran
School of Business and Public Administration, California State University, Bakersfield, California, USA, and
Stephen J. Childe
Plymouth Business School, Plymouth University, Plymouth, UK

Abstract
Purpose – The purpose of this paper is to examine when and how organizations build big data analytics capability (BDAC) to improve supply chain agility (SCA) and gain competitive advantage.
Design/methodology/approach – The authors grounded the theoretical framework in two perspectives: the dynamic capabilities view and contingency theory. To test the research hypotheses, the authors gathered 173 usable responses using a pre-tested questionnaire.
Findings – The results suggest that BDAC has a positive and significant effect on SCA and competitive advantage. Further, the results support the hypothesis that organizational flexibility (OF) has a positive and significant moderation effect on the path joining BDAC and SCA. However, contrary to the belief, the authors found no support for the moderation effect of OF on the path joining BDAC and competitive advantage.
Originality/value – The study makes some useful contributions to the literature on BDAC, SCA, OF, and competitive advantage. Moreover, the results may further motivate future scholars to replicate the findings using longitudinal data.
Keywords Big data, Contingency theory, Dynamic capability view, Analytics capability, Big data analytics capability
Paper type Research paper

1. Introduction
The most successful organizations create supply chains that can respond to sudden and unexpected changes in the market (Lee, 2004). Supply chain management has gained popularity among organizations as a source of competitive advantage (Lee, 2002). Managing supply chains is an extremely challenging task due to the current business outsourcing, globalization, short product life cycle, and continuous advancement in information technology (IT) (Lee, 2002; Christopher and Towill, 2002). Fisher (1997) attempted to match supply chain strategies (i.e. efficient vs responsive) to the product characteristics (i.e. functional vs innovative products). Lee (2002) extended Fisher’s work by proposing four strategies (i.e. efficient/responsive/risk hedging/agile), to accommodate different degrees of demand and supply uncertainty. Hence, according to the situation, the organization’s supply chain agility (SCA) may directly affect its ability to produce and deliver innovative products to their customers at the right time, in the right place, in the right condition and at the right cost (Swafford et al., 2006; Khan and Pillania, 2008). Lee (2002) further argues that agile supply chains utilize strategies aimed at being responsive and flexible to customer’s needs, while the risks of supply shortages or disruptions are hedged by pooling inventory or other capacity resources.

The authors express the sincere thanks to the Editor-in-Chief, the Managing Editor and the Associate Editor. The authors thank ACMA and the McKinsey India for their continued support in data collection. Finally, the authors thank the two anonymous reviewers for their valuable inputs.
However, despite its immense popularity, SCA is still less understood (Braunscheidel and Suresh, 2009). The emerging literature has often used flexibility and agility interchangeably (Braunscheidel and Suresh, 2009). Liu et al. (2013) argues that amidst high environmental uncertainties, organizations are increasingly relying on IT capabilities to gain competitive advantage. Brusset (2016) further argues that supply chain managers are under extreme pressure to improve inventory turnover at minimal cost. Hence, in an effort to adapt the services and goods on offer to meet the customers’ changing tastes and behaviors, supply chain managers are under pressure to build agility in their supply chains to match the intense competition in markets (Gligor et al., 2016). However, an independent stream of research in recent years has led to the conceptualization of SCA as a distinct and different construct from flexibility (Christopher, 2000; Gligor and Holcomb, 2012a, b; Blome et al., 2013; Brusset, 2016; Dubey, Altay, Gunasekaran, Blome, Papadopoulos and Childe, 2018). The literature focusing on SCA and its impact on organizational performance is limited (Gligor et al., 2015) and has focused on the management practices needed to achieve the operational capabilities to enhance SCA from different managerial viewpoints, such as operations, strategy, information systems, marketing, and human resources (Brusset, 2016).

Choi et al. (2017) argue that big data has significant effects on operations management practices. Gunasekaran, Papadopoulos, Dubey, Fosso Wamba, Childe, Hazen and Akter (2017), Gunasekaran, Yusuf, Adeleye and Papadopoulos (2017) further argue that supply chain disruptions have negative effects and agile supply chain enablers were progressively used with the aid of big data and business analytics to achieve better competitive results. Srinivasan and Swink (2017) further argue that although big data analytics has been in use to understand customer intentions/behaviors, the use of analytics for supply chain operational decisions is less understood. Gunasekaran, Papadopoulos, Dubey, Fosso Wamba, Childe, Hazen and Akter (2017), Gunasekaran, Yusuf, Adeleye and Papadopoulos (2017) argue that big data and predictive analytics have positive effects on supply chain performance and organizational performance. Sambamurthy et al. (2003) argue that organizations are increasingly investing in IT capabilities. While some researchers have established the linkage between big data analytics capability (BDAC) and competitive advantage (Akter et al., 2016; Ren et al., 2017; Frisk and Bannister, 2017) and SCA and competitive advantage (Blome et al., 2013; Gligor et al., 2015), little empirical testing of big data analytics and SCA and competitive advantage exists (Sangari and Razmi, 2015; Gunasekaran, Papadopoulos, Dubey, Fosso Wamba, Childe, Hazen and Akter, 2017; Gunasekaran, Yusuf, Adeleye and Papadopoulos, 2017). Hence, we address our first research question:

*RQ1.* What are the distinct and joint effects of BDAC and SCA on competitive advantage?

Boyd et al. (2012) argue that direct effects are crucial, but they seem incapable of explaining real-world complexities. Hence, scholars have acknowledged that the performance effect of supply chain management practices hinges on the environmental context (Sousa and Voss, 2008). Eckstein et al. (2015) argue that such a view is often reflected in contingency theory (CT). The existing research focusing on IT capability and SCA has, however, largely neglected the influence of these contextual factors. In this paper we use a CT perspective to examine under what conditions the BDAC is effective.

Insights derived via big data analysis can provide opportunities for operational improvements (Fosso Wamba et al., 2015; Papadopoulos et al., 2017; Lamba and Singh, 2017; Srinivasan and Swink, 2017; Choi et al., 2017). However, organizations must also convert these valuable insights into actions. Galbraith (2014) argues that the final function to be accelerated by big data is the supply chain. Galbraith (2014, p. 9) further describes how P&G utilizes “decision spheres” where cross-functional teams meet. In case of any emergency plant maintenance, the supply chain managers leverage analytics capability to re-route their trucks and still meet their customers’ demand. The role of organizational flexibility (OF) has
been widely discussed in operations management literature – the ability of an organization to deploy resources quickly and efficiently – as a means to respond to the changing market conditions (Upton, 1994; Swafford et al., 2008; Srinivasan and Swink, 2017). Srinivasan and Swink (2017) argue that analytics capability can provide insights on what to change to match supply and demand; OF enables the organization to determine how to make the appropriate changes. However, such crucial effects have not been addressed by prior research theoretically or subjected to empirical testing. Thus, we specify our second research question as follows:

*RQ2. What are the effects of OF on the relationships between big data analytics and SCA/competitive advantage?*

We provide answers to our research questions, using data from 173 survey responses from experienced supply chain managers. To theoretically substantiate our empirical results, we have integrated the dynamic capabilities view (DCV) (Teece et al., 1997) and CT, because neither perspective can, on its own, explain the direct impacts of BDAC on SCA and competitive advantage and the contextual conditions under which they are effective.

The paper is organized as follows. In Section 2 of the paper, we present the underpinning theories. In Section 3, we illustrate our theoretical framework and outline our research hypotheses accordingly. In Section 4, we present our research design, including discussion of instrument development, sampling design, and data collection. In Section 5, we present our statistical analyses and results. In Section 6, we provide discussion of our statistical results. In Section 7, we conclude our study with theoretical contribution and managerial implications, limitations and further research directions.

### 2. Underpinning theories

#### 2.1 DCV

Following criticism of the resource-based view, which often fails to provide explanation of how and in what context the resources can provide competitive advantage to a firm (Eisenhardt and Martin, 2000), scholars have argued that the DCV provides explanation for the organization’s competitive advantage in changing environments (Teece et al., 1997; Sirmon et al., 2010; Eisenhardt and Martin, 2000; Singh et al., 2013). Teece et al. (1997, p. 516) defined dynamic capabilities as, “the firm’s ability to integrate, build and reconfigure internal and external competencies to address rapidly changing environments.” Teece et al. (1997) further argue that dynamic capabilities include the capabilities to sense and shape opportunities, to seize opportunities, and to maintain competitiveness through enhancing, combining, protecting and reconfiguring a firm’s resources. Within the context of a highly uncertain environment, dynamic capabilities are simple, experiential, unstable processes that rely on rapidly created new insights that enable combination, transformation, or renewal of resources and competencies into capabilities which are essential for uncertain markets (Eisenhardt and Martin, 2000; Eckstein et al., 2015). Based on these arguments scholars have considered big data analytics as a dynamic capability (Fosso Wamba et al., 2017) that results from the organization’s ability to reconfigure firm-level resources.

#### 2.2 BDAC

There is increasing debate about the importance of big data analytics in supporting the strategic goals of an organization (Davenport, 2006; Manyika et al., 2011; Prescott, 2014; Mishra et al., 2016, 2017; Roden et al., 2017; Ren et al., 2017; Choi et al., 2017; Fosso Wamba, 2017; Jabbour et al., 2017), but there is as yet no consensus about how best to organize big data analytics efforts within the organization and what core analytics processes the organization should support (Galbraith, 2014). Following Manyika et al. (2011), we argue that big data are data whose volume, velocity, and variety make it difficult for an
organization to manage, analyze, and extract valuable insights using conventional and traditional methods. The term analytics refers to “the process that extracts valuable insights from data via creating and distributing reports, building and deploying statistical and data-mining models, exploring and visualizing data, sense-making and other related techniques” (Grossman and Siegel, 2014, p. 20). Hence, we can argue that BDAC is an organizational facility with tools, techniques, and processes that enable the organization to process, visualize, and analyze data, thereby producing insights that enable data-driven operational planning, decision making and execution (Srinivasan and Swink, 2017). In the context of supply chain management, BDAC enables to firms to examine alternatives related to supply and demand uncertainties (Waller and Fawcett, 2013; Hazen et al., 2014; Wang et al., 2016).

2.3 SCA

Lee (2004) argues that organizations are increasingly investing in building agility in supply chains to respond to sudden and unexpected changes in the market. Swafford et al. (2006) argue that SCA affects the ability of an organization to produce and deliver innovative products to their customers in a timely and cost-effective manner. Braunscheidel and Suresh (2009) noted that with intense competitive pressures as well as high levels of turbulence and uncertainty, organizations require agility in their supply chains. The agility in supply chains provides superior value as well as overcoming disruption risks and ensuring uninterrupted service to customers (Braunscheidel and Suresh, 2009; Blome et al., 2013; Gligor et al., 2016; Brusset, 2016). Christopher (2000) has noted that numbers of characteristics that a supply chain must possess to be agile are:

- Market sensitive – it closely monitor the changes in demand pattern.
- Virtual – information sharing among partners in supply chain is critical.
- Network-based – helps to build flexibility in supply chain network.
- Process integration – it has a high degree of process interconnectivity between the network members.

Hence, these characteristics help the organizations to meet customer demand by providing the right product at the right time, place, and price. Some notable examples include Dell, Wall-Mart and Amazon (Lee, 2004). Swafford et al. (2006) found that organizations’ SCA is directly and positively impacted by flexibility in manufacturing and procurement/sourcing processes. Braunscheidel and Suresh (2009) observed, based on empirical results, that besides flexibility, internal cross-functional integration and external integration with key customers and suppliers are crucial for enhancing agility in supply chains. Eckstein et al. (2015) view SCA as a dynamic capability that not only helps to meet customers’ demand but also helps to enhance the firm’s profitability. Dubey, Altay, Gunasekaran, Blome, Papadopoulos and Childe (2018) further noted that supply chain visibility enhances SCA via bundling organizational resources (i.e. data connectivity and information sharing). Hence, we can argue that agility is a desired property of a supply chain that enables it to respond to short-term changes in demand and supply quickly and handle external disruptions smoothly.

2.4 OF

Volberda (1996, p. 361) defines OF, as “the degree to which an organization has a variety of managerial capabilities and the speed at which they can be activated, to increase the control capacity of the management and improve the controllability of the organization.” Hence, we argue that OF is the organizational ability that enables organizations to operate in a
turbulent environment (Braunscheidel and Suresh, 2009; Sharma et al., 2010; Srinivasan and Swink, 2017). Sanchez (1993) argues that in dynamic environments, an organization can gain competitive advantage by creating strategic flexibility. Sanchez (1995) further argues that flexibility is constrained not only by resources but also by the ways a firm uses the resources (see also Upton, 1994; Suarez et al., 1996; Sanchez, 1997; Sanchez and Mahoney, 1997; Liu et al., 2009).

2.5 Competitive advantage

Porter (1985) argues that firms can gain competitive advantage by identifying and implementing generic strategies and addresses the interplay between types of competitive advantage – cost and differentiation – and the scope of the firm’s activities. Barney (1991, p. 102) defined competitive advantage, “a firm is said to have competitive advantage when it is implementing value creating strategy not simultaneously being implemented by any current or potential competitors.” Peteraf (1993) argues that competitive advantage is the ability of an organization to maintain or sustain above-normal returns. However, Peteraf (1993) further argues that there are four cornerstones of competitive advantage: heterogeneity, ex post limits to competition, imperfect mobility and ex ante limits to competition. Barney (1991) argues that an organization can derive competitive advantage by creating bundles of strategic resources and/or capabilities. Reed and DeFillippi (1990) argue that competitive advantage can be derived from numerous sources. For instance, competencies are within the organization’s control and can be exploited to generate competitive advantage for superior performance (Hinterhuber, 2013).

3. Theoretical framework and hypotheses development

The foundation of our theoretical model (see Figure 1) comprises two pillars: DCV and CT. Due to rapidly changing environments, the DCV has gained immense popularity among management scholars who seek to investigate how the bundling of firm resources and competencies can provide competitive advantage to a firm operating in a highly uncertain environment. Consistent with the previous arguments, information processing capability is seen a solution for uncertainty. The need for BDAC is further heightened by volatile and complex task environments, where high levels of uncertainty make effective planning and decision making difficult. Using the logic of fit, scholars have argued that organizational

Figure 1.
Theoretical framework
flexibilities of several types are more valuable in highly uncertain environments (Swamidass and Newell, 1987; Pagell and Krause, 1999). Hence, we first draw direct links from BDAC to connect SCA and CA. Second, in order to examine the interaction effects of OF on the direct effects of BDAC on SCA and CA we draw links on the paths joining BDAC to SCA and CA. To further examine the direct effect of SCA on CA, we draw a link joining SCA and CA and propose four research hypotheses grounded in DCV and CT. These hypotheses do not exclude the possibility that other factors may influence the effect of BDAC on SCA/CA; we have controlled for these factors during our model testing and in our subsequent discussion.

3.1 BDAC and SCA
Swafford et al. (2008) found that IT capability has positive effect on SCA. Srinivasan and Swink (2017) noted that supply chain visibility is a prerequisite for building data analytics capability and vice versa. Supply chain visibility and BDACs are complementary, in the sense that each supports the other (Gunasekaran, Yusuf, Adeleye and Papadopoulos, 2017; Gunasekaran, Papadopoulos, Dubey, Fosso Wamba, Childe, Hazen and Akter, 2017). Supply chain visibility is a desired organizational capability to mitigate risk resulting from supply chain disruptions (Jüttner and Maklan, 2011). Following Srinivasan and Swink’s (2017) arguments that organizations investing in building supply chain visibility capability are likely to invest in BDAC, Dubey, Altay, Gunasekaran, Blome, Papadopoulos and Childe (2018) found a positive impact of supply chain visibility on SCA. Hence, we may argue that the use of data technology may help managers to sense the rapid changes in environments, so they can develop business continuity plans that may help to quickly respond to the changes. Thus:

H1. Big data analytics has a positive impact on SCA.

3.2 BDAC and competitive advantage
Competitive advantage refers to any advantage that a firm has over their competitors (Porter, 1985). Chen et al. (2012) argue that big data presents an immense opportunity to achieve competitive advantage. LaValle et al. (2011) noted that top-performing organizations use analytics five times more than low performers. Raffoni et al. (2018) argue that big data analytics, if used cautiously, can help the organization to achieve better performance. Akter et al. (2016) argue that BDAC has a positive impact on organizational performance. Kache and Seuring (2017) argue that the use of big data analytics is still in its infancy stage, but despite challenges big data analytics appears to offer immense opportunities. Zhang et al. (2017) argue that organizations are increasingly exploiting big data to improve organizational competitiveness. Gunasekaran, Yusuf, Adeleye and Papadopoulos (2017), Gunasekaran, Papadopoulos, Dubey, Fosso Wamba, Childe, Hazen and Akter (2017) noted that the big data and predictive analytics capability have positive impact on supply chain and organizational performance. Hence, we hypothesize:

H2. Data analytics has positive impact on organization’s competitive advantage.

3.3 SCA and competitive advantage
The great companies create agile supply chains to respond to sudden and unexpected changes in markets. Brusset (2016) argues that SCA is desirable as in most industries, both demand and supply fluctuate rapidly. Lee (2004) argues that organizations like H&M, Mano and Zara use SCA to differentiate themselves from their competitors. Whitten et al. (2012) tested empirically, using a survey of 132 respondents, that SCA along with other capabilities (i.e. supply chain adaptability and supply chain alignment) has a positive impact on
supply chain performance and supply chain performance further positively affects organizational performance. Gligor et al. (2015) tested using a survey of 283 that SCA has a positive impact on financial performance under the mediating effects of customer effectiveness and cost efficiency. Thus, we test:

\[ H3. \text{SCA has a positive impact on competitive advantage.} \]

3.4 Moderating effects of OF

The BDAC provides useful insights based on processing of the data gathered from multiple sources (Srinivasan and Swink, 2017; Choi et al., 2017). Galbraith (1973, 1974) noted that organizations need flexibility to implement decisions quickly and efficiently, especially decisions that span various functions. OF has been noted as one of the key levers to reduce supply chain risk (Braunschweig and Suresh, 2009). Hence, we posit that organizations can more effectively take advantage of new insights gained from BDAC when they possess high levels of OF. Organizations with better OF are more capable of coping with demand and supply uncertainties (Swafford et al., 2006) and gain competitive advantage (Yusuf et al., 2004, 2014). Consequently, those organizations have better capabilities to improve SCA than those organizations which rely on decisions based on limited data sets or mechanistic models of processing data. Thus:

\[ H4. \text{OF positively moderates the relationship between BDAC and: (a) SCA and (b) competitive advantage.} \]

4. Research design

4.1 Sample and data collection

We analyzed data collected in 2015 through a survey of Automotive Components Manufacturers Association (ACMA of India), to test our theoretical model. ACMA and McKinsey administered this cross-sectional survey. The unit of analysis is the organization and the survey was developed for a single respondent. Our research team sent e-mail invitations to 745 supply chain managers drawn from the ACMA database. The sampling frame included senior supply chain managers from auto components manufacturing organizations located in various parts of India. For this study, senior supply chain managers in logistics, production, procurement, and information systems functions were targeted, as they are likely to have relevant information related to materials and information flow as well as supply chain innovation initiatives. With regard to the four supply chain types (Lee, 2002), it should be noted that all the respondents were in the auto components industry and could expect to be broadly comparable.

Overall, we received 173 complete and usable responses, representing an effective response rate of 23.22 percent (see Table I). In this study, we eliminated those respondents whose titles were not directly related to supply chain functions and had less experience. The resulting sample held senior managerial positions such as vice president, general manager, CXO (C-suite managers), director, head, senior manager and manager with more than 15 years of experience. We also included responses from analysts and planners. We provide the profile of the responding organizations in Table I and the profile of the respondents in Table AI.

In survey-based research, there is always a potential for biases. As we used a survey-based approach to gather data, we tested for non-response bias through comparison of early responses and late responses, following Armstrong and Overton’s (1977) suggestions. The t-tests yielded no statistically significant differences between early and late responses, indicating that non-response bias is not a problem in our study.
4.2 Measures
Following Malhotra and Grover’s (1998) suggestions, we used established scales from literature in our study. This was feasible for measures of data analytics, OF, SCA, and competitive advantage. We made minor modifications in the wording of our questionnaire based on pre-tests to improve the performance of the questionnaire. All the scales were designed in five-point Likert format with anchors 1 = strongly disagree and 5 = strongly agree.

In addition, we identified three control variables, which may influence the exogenous and endogenous variables and may cause unwanted sources of variance. First, firm size, following Cao and Zhang’s (2011) arguments that smaller firms have fewer resources for the implementation of supply chain management practices, and Wagner and Neshat (2012) who noted that larger organizations are more vulnerable to supply chain disruptions. The number of employees (logarithmic) measured the size of the firm (Gligor et al., 2015).

Second, we included industry dynamism in order to level out the effects of uncertainty across industry segments. Aldrich (1979) argues that dynamism reflects the unpredictability and volatility, of the changes in the industry that heightens the uncertainty of the organizations’ predictions. We measured industry dynamism on a five-point Likert format anchored as, 1 = slow and 5 = rapid, with items reflecting the industry rates of change for product/service introduction, operating processes, customer tastes/preferences, and research and development (Brandon-Jones et al., 2014). Third, we controlled for the age of the organization. Gligor et al. (2015) argue that the age of the organization can influence the implementation of supply chain practice and therefore, impact upon competitive advantage. The firm age is calculated as the number of years since the firm foundation (logarithmic) (see, Srinivasan and Swink, 2017). Table II shows the summary of the items used for measures.

5. Data analyses and results
We used WarpPLS 5.0 software to test our model. The software employs the partial least squares (PLS) structural equation modeling method or in short form PLS SEM (Kock, 2014, 2015). PLS is a prediction-oriented tool which allows researchers to assess the predictive validity of the exogenous variables (Peng and Lai, 2012). Scholars argue that PLS is better suited for explaining complex relationships as it avoids two serious problems: inadmissible solutions and factor indeterminacy (see, Peng and Lai, 2012;
<table>
<thead>
<tr>
<th>Construct</th>
<th>Reference</th>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big data analytics</td>
<td>Akter et al. (2016), Srinivasan and Swink (2017)</td>
<td>BDAC1</td>
<td>We use advanced tools (like optimization/regression/simulation) for data analysis</td>
</tr>
<tr>
<td>capability (BDAC)</td>
<td></td>
<td>BDAC2</td>
<td>We use data gathered from multiple sources (like company reports, tweets, Instagram, YouTube) for data analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BDAC3</td>
<td>We use data visualization techniques to assist decision makers in understanding complex information extracted from large data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BDAC4</td>
<td>Our dashboards display information, which is useful for carrying out necessary diagnosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BDAC5</td>
<td>We have connected dashboard applications or information with the manager’s communication devices</td>
</tr>
<tr>
<td>Organizational flexibility (OF)</td>
<td>Sethi and Sethi (1990), Upton (1994)</td>
<td>OF1</td>
<td>We can quickly change organizational structure to respond to demand and supply uncertainties</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OF2</td>
<td>Our organization can cost effectively respond to sudden changes in the market</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OF3</td>
<td>Our organization is more flexible than our competitors in changing our organizational structure</td>
</tr>
<tr>
<td>Supply chain agility (SCA)</td>
<td>Gligor et al. (2015)</td>
<td>SCA1</td>
<td>Our organization can quickly detect changes in our environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCA2</td>
<td>Our organization can quickly identify opportunities in its environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCA3</td>
<td>Our organization can quickly sense threats in its environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCA4</td>
<td>Our organization continuously collects information from suppliers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCA5</td>
<td>Our organization continuously collects information from customers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCA6</td>
<td>We make quick decisions to deal with changes in environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCA7</td>
<td>When needed we can adjust our supply chain operations to the extent necessary to execute our decisions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCA8</td>
<td>Our organization can increase its short-term capacity as needed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCA9</td>
<td>We can adjust the specification of orders as requested by our customers</td>
</tr>
<tr>
<td>Competitive advantage (CA)</td>
<td>Tracey et al. (1999), Vorhies and Morgan (2005)</td>
<td>CA1</td>
<td>Our customers are satisfied with our product quality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA2</td>
<td>We deliver value to our customer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA3</td>
<td>We deliver at the right time what our customers want</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA4</td>
<td>Our market share growth is significant in comparison to our customers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA5</td>
<td>We are able to acquire new customers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA6</td>
<td>We have reached our financial goals</td>
</tr>
<tr>
<td>Industry dynamism (ID)</td>
<td>Brandon-Jones et al. (2014)</td>
<td>ID1</td>
<td>Our product and services become outdated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ID2</td>
<td>Our organization continuously introduces new products and services</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ID3</td>
<td>Our organization introduces new operating processes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ID4</td>
<td>The customers taste and preferences in our industry changes fast</td>
</tr>
<tr>
<td>Age of the organization (OA)</td>
<td>Gligor et al. (2015)</td>
<td>OA</td>
<td>Logarithmic value of number of years</td>
</tr>
<tr>
<td>Organization size (OS)</td>
<td>Gligor et al. (2015)</td>
<td>OS</td>
<td>Logarithmic value of number of employees</td>
</tr>
</tbody>
</table>

**Table II.** Measures

---

*MD 57,8*
Henseler et al., 2014; Moshtari, 2016; Pratono, 2016; Akter et al., 2017; Martí-Ballester and Simon, 2017; Dubey, Gunasekaran, Childe and Papadopoulos, 2018). In our study, we aim to examine the prediction or explanatory power of BDAC. The relationships between two variables – BDAC and SCA – are not examined in literature. With no theoretical foundation that explains the relationships between these two variables, PLS becomes the most suitable technique for data analysis (Peng and Lai, 2012). We carried out our model estimation based on Peng and Lai’s (2012) suggestions in two stages: examining the reliability and validity of the measurement model and then analyzing the structural model.

5.1 Measurement model
A series of procedures were used to determine convergent and discriminant validity for the constructs used in our model (see Figure 1). In support of convergent validity, we noted that factor loadings were significant except for a few items which we dropped from our study. Next, we found that the average variance extracted (AVE) of each construct was greater than 0.7. As shown in Table AII, the loadings are in an acceptable range and they are significant at the 0.01 level (Fornell and Larcker, 1981). The discriminant validity was assessed via AVE comparisons (see Table III). The square roots of the AVEs were greater than all of the inter-construct correlations; it is a strong evidence of sufficient discriminant validity.

5.2 Common method bias (CMB)
Ketokivi and Schroeder (2004) argue that data gathered using a survey-based instrument from a single source has potential biases. Podsakoff et al. (2003) argue that in the case of self-reported data, there is potential for CMBs resulting from multiple sources such as consistency motif and social desirability. Hence, we designed our survey to minimize the CMB effect using different scale formats and anchors for independent, moderating, and dependent variables. In addition, we performed several statistical analyses to assess the extent of CMB. First, following Podsakoff and Organ (1986) we performed a conservative version of Harman’s one-factor test. The results from this test showed that the single factor explains 43.69 percent (approx.), of total variance, demonstrating that CMB is not a significant concern. Second, we tested for CMB using the marker technique (Lindell and Whitney, 2001). We used an unrelated variable to partial out the correlations caused by CMB. We also calculated the significant value of the correlations based on Lindell and Whitney’s (2001) equations. We noted no significant differences between adjusted and unadjusted correlations. Based on these results we consider that the potential effects of common method variance are negligible.

Guide and Ketokivi (2015) noted that causality is an important issue that should be examined prior to hypothesis testing. Following Kock’s (2015) suggestions, we calculated the nonlinear bivariate causality direction ratio (NLBCDR). The NLBCDR refers to “an interesting property of nonlinear algorithms [...] that bivariate nonlinear coefficients of

<table>
<thead>
<tr>
<th></th>
<th>BDAC</th>
<th>OF</th>
<th>CA</th>
<th>SCA</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDAC</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OF</td>
<td>0.24</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>0.24</td>
<td>0.56</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCA</td>
<td>0.61</td>
<td>0.37</td>
<td>0.41</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>0.08</td>
<td>0.01</td>
<td>0.17</td>
<td>0.20</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Table III. Correlations among major constructs

Notes: BDAC, big data analytics capability; OF, organizational flexibility; SCA, supply chain agility; CA, competitive advantage; ID, industry dynamism. (AVE) are in italic
association vary depending upon the hypothesized direction of the causality. That is, they tend to be stronger in one direction than the other, which means that the residual (or error) is greater when the hypothesized direction of causality is in one way or the other. Hence, the NLBCDR index is a measure of the extent to which bivariate nonlinear coefficients of association provide support for the hypothesized directions of the causal links in the model” (Kock, 2015, pp. 52-53). The desired acceptable value is greater than 0.7. In our model NLBCDR = 0.818, which is greater than the cut-off value. Hence, based on these results we can argue that endogeneity is not a serious concern in our study. We further tested the model fit and quality indices (see Table AIII).

5.3 Hypothesis testing
Figure 2 presents the estimates obtained via PLS SEM analysis. The model explains a significant amount of variance for SCA \( (R^2 = 0.29) \) and competitive advantage \( (R^2 = 0.69) \). We have reported the PLS path coefficients and the corresponding \( p \)-values for the model in Table IV (H1-H3) and Table V (H4a and H4b). The links BDAC → SCA \( (\beta = 0.32, p < 0.01) \),

![Figure 2. Causal model](image)

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Effect of</th>
<th>Effect on</th>
<th>( \beta )</th>
<th>( p )</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>BDAC</td>
<td>SCA</td>
<td>0.32</td>
<td>***</td>
<td>Supported</td>
</tr>
<tr>
<td>H2</td>
<td>BDAC</td>
<td>CA</td>
<td>0.28</td>
<td>***</td>
<td>Supported</td>
</tr>
<tr>
<td>H3</td>
<td>SCA</td>
<td>CA</td>
<td>0.41</td>
<td>***</td>
<td>Supported</td>
</tr>
</tbody>
</table>

**Control variables**

<table>
<thead>
<tr>
<th></th>
<th>Effect of</th>
<th>Effect on</th>
<th>( \beta )</th>
<th>( p )</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>CA</td>
<td>0.81</td>
<td>***</td>
<td></td>
<td>Significant</td>
</tr>
<tr>
<td>OA</td>
<td>CA</td>
<td>0.05</td>
<td>*</td>
<td></td>
<td>Not significant</td>
</tr>
<tr>
<td>OS</td>
<td>CA</td>
<td>0.00</td>
<td>*</td>
<td></td>
<td>Not significant</td>
</tr>
<tr>
<td>ID</td>
<td>SCA</td>
<td>0.42</td>
<td>***</td>
<td></td>
<td>Significant</td>
</tr>
<tr>
<td>OA</td>
<td>SCA</td>
<td>0.1</td>
<td>*</td>
<td></td>
<td>Not significant</td>
</tr>
<tr>
<td>OS</td>
<td>SCA</td>
<td>0.03</td>
<td>*</td>
<td></td>
<td>Not significant</td>
</tr>
</tbody>
</table>

**Table IV. Structural estimates (H1-H3)**

**Notes:** BDAC, big data analytics capability; SCA, supply chain agility; CA, competitive advantage; ID, industry dynamism; OA, age of the organization; OS, organizational size. *\( p > 0.1 \); ***\( p < 0.01 \)
BDAC→CA ($\beta = 0.28$, $p < 0.01$), and SCA→CA ($\beta = 0.41$, $p < 0.01$) are positively related. Thus, we can argue, based on $\beta$ values and their corresponding $p$-values, that $H1$-$H3$ were supported. The control variables organization age and organizational size (OS) do not have significant effect in this model (see Table IV). However, industry dynamism has a significant effect on SCA and CA.

Next, our hypothesis $H4$ was tested for the moderation effect of OF on the path connecting data analytics capability and supply chain resilience ($H4a$) and data analytics capability and competitive advantage ($H4b$). $H4a$ ($\beta = 0.28$, $p < 0.01$) was found to be supported (see Table IV). However, $H4b$ ($\beta = 0.09$, $p > 0.1$) was not supported.

Next, we examined the explanatory power of our proposed theoretical model. For this, we examined the explanatory power ($R^2$) of the endogenous construct. The $R^2$ for SCA is 0.29 which is moderately strong and for CA is 0.69 which is strong (Chin, 1998) (see Figure 2). We further examined the $f^2$ value of the BDAC using Cohen’s $f^2$ formula. Consequently, the effect size of BDAC on SCA is 0.002 and on CA is 0.003 (see Table VI) which were greater than the cut-off value of zero. Next, we examined the model’s capability to predict. Stone-Geiser’s $Q^2$ values for the endogenous constructs were SCA (0.3) and CA (0.66) (see Table VI) for BDAC which is greater than zero, indicating acceptable predictive relevance (Peng and Lai, 2012).

6. Discussion

Our results provide a better understanding of the impact of BDAC on SCA and competitive advantage, answering the calls for research (Brusset, 2016; Gunasekaran, Yusuf, Adeleye and Papadopoulos, 2017; Gunasekaran, Papadopoulos, Dubey, Fosso Wamba, Childe, Hazen and Akter, 2017; Dubey, Altay, Gunasekaran, Blome, Papadopoulos and Childe, 2018). Our results contribute to building and refining theories of BDAC, SCA, OF, and competitive advantage and provide empirically grounded normative suggestions to management practitioners. The results provide improve our understanding about the relationship between BDAC, SCA, and competitive advantage. Our study is the first to offer a rigorous empirical test of the distinct effects of BDAC on SCA and competitive advantage, which was called for in previous research (Gunasekaran, Yusuf, Adeleye and Papadopoulos, 2017; Gunasekaran, Papadopoulos, Dubey, Fosso Wamba, Childe, Hazen and Akter, 2017). We found that age of the firm (OA) and OS have no significant effect on SCA and competitive advantage. We interpret these observations as evidence the auto components

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Effect of</th>
<th>Effect on</th>
<th>$\beta$</th>
<th>$p$</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H4a$</td>
<td>BDAC×OF</td>
<td>SCA</td>
<td>0.28</td>
<td>***</td>
<td>Supported</td>
</tr>
<tr>
<td>$H4b$</td>
<td>BDAC×OF</td>
<td>CA</td>
<td>0.09</td>
<td>*</td>
<td>Not supported</td>
</tr>
</tbody>
</table>

Notes: BDAC, data analytics capability; SCA, supply chain agility; CA, competitive advantage; OF, organizational flexibility. *$p > 0.1$; ***$p < 0.01$

<table>
<thead>
<tr>
<th>Construct</th>
<th>$R^2$</th>
<th>$Q^2$</th>
<th>$f^2$ in relation to</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDAC</td>
<td>–</td>
<td>–</td>
<td>0.002</td>
</tr>
<tr>
<td>OF</td>
<td>–</td>
<td>–</td>
<td>0.003</td>
</tr>
<tr>
<td>SCA</td>
<td>0.29</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>0.69</td>
<td>0.66</td>
<td></td>
</tr>
</tbody>
</table>

Table V. Structural estimates ($H4a$ and $H4b$)

Table VI. $R^2$, prediction and effect size
sector is increasingly using big data analytics to improve SCA and gain competitive advantage in the face of rapid change, and the size and age of the firm play little role in the BDAC-SCA-competitive advantage relationship. However, industry dynamism has positive effects on SCA and competitive advantage. We interpret these observations to suggest that dynamic elements such as the rate with which new products or services are introduced, operating processes, customer tastes/preferences, and research and development may have a major role to play in the big data analytics-SCA-competitive advantage relationship. While this may be contrary to our expectations, the role of industry dynamism may be further explored in the context of the big data analytics-SCA-competitive advantage relationship.

Our hypotheses that under high OF the relationship between BDAC and SCA/competitive advantage will be strengthened were partially supported. Specifically, our results support our belief that OF positively moderates the link joining BDAC and SCA. However, our expectation that OF would positively moderate the link from BDAC to competitive advantage, this was not supported. These results may be due to the inclusion of OF as a key contextual factor in our theoretical model. Hence, we suggest further research is needed to examine whether OF has a role to play on the impact of BDAC and competitive advantage.

7. Conclusions
7.1 Implications for theory
Brusset (2016) found empirically that visibility may not have a significant effect on SCA. Contrary to Brusset’s (2016) findings, Dubey, Altay, Gunasekaran, Blome, Papadopoulos and Childe (2018) observed that supply chain visibility has a positive and significant effect on SCA. Srinivasan and Swink (2017) argue that visibility is created via external relations, which may help decision makers to sense changes in customer and competitors markets, including demands, pricing and promotional actions, and product inventories. Hence, the organizations that develop demand and supply visibility are also better positioned to develop and deploy systems and processes supporting analytics capability. Building upon this tautology, we posited that organizations that seek to enhance supply chain visibility would invest in building BDAC, which will help them handle large data derived from various sources to extract useful insights. Building on Srinivasan and Swink’s (2017) arguments, we tested the direct impact of BDAC on SCA. Our study is the first to provide an empirical test of the distinct effects of BDAC on SCA and competitive advantage. The information derived via BDAC provides firms with real-time information regarding changes in future product demand due to changes in downstream inventories, promotions, and sales. Moreover, supplier-sourced data provide information regarding supply shortages and excess inventories resulting from changes in upstream inventories, capacities, and the status of orders and shipments. We have thus answered the research calls of prior literature (see Gunasekaran, Yusuf, Adeleye and Papadopoulos, 2017; Gunasekaran, Papadopoulos, Dubey, Fosso Wamba, Childe, Hazen and Akter, 2017; Dubey, Altay, Gunasekaran, Blome, Papadopoulos and Childe, 2018). Moreover, our study is the first to empirically investigate the moderating effect of OF on the paths connecting BDAC and SCA/competitive advantage. Thus, we contribute to the literature by addressing the need for more holistic understanding of distinct relationships among contingencies (i.e. OF), response alternatives (i.e. BDAC) and multiple performance outcomes (i.e. SCA and competitive advantage). In doing so, we contribute to our understanding of the specific contextual factors under which BDAC can effectively improve SCA. Hence, by integrating the perspectives of the DCV and CT, we provide a solid theoretical grounding for our empirical investigation of CT as a complement to DCV, given its shortcomings in recognizing the complexity involved while bundling resources and capabilities (Eckstein et al., 2015).
7.2 Implications for practice
By empirically testing our theoretical model we provide established evidence (to surpass anecdotal evidence) that organizations in our sample do indeed benefit from BDAC to sense market changes, including demands, pricing and promotional schemes of their competitors, and product inventories. Moreover, our results strengthen the notion that managers who can exploit the innovative technology in attempts to build capability at supply chain level can expect their organizations to gain competitive edge over their competitors (Davenport, 2006; Akter et al., 2016; Srinivasan and Swink, 2017). However, before building BDAC, a thorough understanding of OF is critical. OF stems from knowledge and abilities to change organizational structures and resource allocations quickly and efficiently. Our results suggest that OF serves as a complementary capability to BDAC in consistently improving SCA in highly uncertain environments. Even though research provides evidence to show management practitioners that BDAC does indeed pay off, practitioners require a better understanding of how they can develop this critical dynamic capability. This issue is of critical importance, as BDAC requires a significant investment of resources and effort. Hence, without appropriate understanding of the resources and the competences needed to build BDAC, practitioners may not achieve the desired outcome via BDAC.

7.3 Limitations and further research opportunities
Like any other studies, the results of our study are subject to several limitations that must be taken into consideration while interpreting these results. First, we tested our theoretical model using data gathered at one point in time. Following Guide and Ketokivi (2015), we can argue that cross-sectional data are one of the major causes of CMB and causality. While we have performed several statistical analyses which may provide evidence that CMB and causality were not a major concern in our study, a longitudinal design would help to reduce the possibility of CMB that undermines the validity of studies with data from a single source at a single point in time.

Second, we used the perspectives of DCV and CT. However, future studies may use other theory or theories to provide better explanation. Scholars may investigate other organizational capabilities or assets considered as complements to BDAC. Moreover, other theoretical perspectives including knowledge based view, absorptive capacity, organizational learning, organizational culture and top management commitment might provide useful extensions to our study.

Third, we believe that survey-based research has its own limitations. Hence, future researchers may be able to address some unanswered questions via case based methods.

Finally, the demographic of our research sample may limit the generalizability of our findings. Thus, the research findings should be applied to other contexts with caution. We acknowledge that any study using a survey-based approach often faces a generalizability issue. It is very difficult to obtain a sample that could claim to be truly representative of the whole population. Still, future research should be conducted over a period with samples from more industries, countries, and informants with diverse backgrounds.

References


### Appendix 1

#### Table AI.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Category</th>
<th>Respondents</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>163</td>
<td>94.22</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>10</td>
<td>5.78</td>
</tr>
<tr>
<td>Age</td>
<td>&gt; 50</td>
<td>110</td>
<td>63.58</td>
</tr>
<tr>
<td></td>
<td>40-50</td>
<td>50</td>
<td>28.90</td>
</tr>
<tr>
<td></td>
<td>30-39</td>
<td>13</td>
<td>7.51</td>
</tr>
<tr>
<td></td>
<td>&lt; 30</td>
<td>Nil</td>
<td>Nil</td>
</tr>
<tr>
<td>Education level</td>
<td>Doctorate</td>
<td>6</td>
<td>3.47</td>
</tr>
<tr>
<td></td>
<td>Postgraduate degree in management</td>
<td>60</td>
<td>34.68</td>
</tr>
<tr>
<td></td>
<td>Postgraduate degree in engineering</td>
<td>70</td>
<td>40.46</td>
</tr>
<tr>
<td></td>
<td>Postgraduate degree in science</td>
<td>12</td>
<td>6.94</td>
</tr>
<tr>
<td></td>
<td>Bachelor’s degree</td>
<td>23</td>
<td>13.29</td>
</tr>
<tr>
<td></td>
<td>Associate degree</td>
<td>2</td>
<td>1.16</td>
</tr>
</tbody>
</table>

#### Table AII.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Items</th>
<th>Factor loadings</th>
<th>Variance</th>
<th>Error</th>
<th>SCR</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDAC</td>
<td>BDAC1</td>
<td>0.75</td>
<td>0.57</td>
<td>0.43</td>
<td>0.80</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>BDAC2</td>
<td>0.66</td>
<td>0.44</td>
<td>0.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BDAC3</td>
<td>0.86</td>
<td>0.74</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OF</td>
<td>OF1</td>
<td>0.95</td>
<td>0.90</td>
<td>0.10</td>
<td>0.97</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>OF2</td>
<td>0.95</td>
<td>0.91</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OF3</td>
<td>0.96</td>
<td>0.92</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>CA2</td>
<td>0.96</td>
<td>0.92</td>
<td>0.08</td>
<td>0.98</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>CA3</td>
<td>0.97</td>
<td>0.94</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CA4</td>
<td>0.94</td>
<td>0.87</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CA5</td>
<td>0.94</td>
<td>0.89</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CA6</td>
<td>0.95</td>
<td>0.91</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCA</td>
<td>SCA2</td>
<td>0.95</td>
<td>0.91</td>
<td>0.09</td>
<td>0.97</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>SCA3</td>
<td>0.93</td>
<td>0.87</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCA4</td>
<td>0.93</td>
<td>0.87</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCA5</td>
<td>0.95</td>
<td>0.90</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCA6</td>
<td>0.90</td>
<td>0.81</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCA7</td>
<td>0.88</td>
<td>0.78</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>ID1</td>
<td>0.78</td>
<td>0.61</td>
<td>0.39</td>
<td>0.93</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>ID2</td>
<td>0.96</td>
<td>0.92</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ID3</td>
<td>0.96</td>
<td>0.92</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ID4</td>
<td>0.81</td>
<td>0.65</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** BDAC, big data analytics capability; OF, organizational flexibility; CA, competitive advantage; SCA, supply chain agility; ID, industry dynamism; (BDAC1, CA1, SCA1, SCA8, SCA9, SCA10 were dropped from our analysis due to weak loadings)
Table AIII. Model fit and quality indices

<table>
<thead>
<tr>
<th>Model fit and quality indices</th>
<th>Value from analysis</th>
<th>Acceptable if</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>0.183, (p = 0.003)</td>
<td>(p &lt; 0.05)</td>
<td>Rosenthal and Rosnow (1991)</td>
</tr>
<tr>
<td>ARS</td>
<td>0.489, (p &lt; 0.001)</td>
<td>(p &lt; 0.05)</td>
<td></td>
</tr>
<tr>
<td>AVIF</td>
<td>1.134, (p &lt; 0.001)</td>
<td>(p &lt; 0.05)</td>
<td>Kock (2015)</td>
</tr>
<tr>
<td>Tenenhaus GoF</td>
<td>0.693</td>
<td>Large if (\geq 0.36)</td>
<td>Tenenhaus et al. (2005)</td>
</tr>
</tbody>
</table>

Corresponding author
Rameshwar Dubey can be contacted at: r.dubey@montpellier-bs.com
Big data management: implications of dynamic capabilities and data incubator

S.M. Riad Shams
Graduate School of Economics and Management,
Ural Federal University, Yekaterinburg, Russian Federation, and
Ludovico Solima
Department of Business, University Della Campania “L. Vanvitelli”, Napoli, Italy

Abstract

Purpose – Big data management research and practice, however, have received enormous interest from academia and industry; the extant literature demonstrates that the authors have limited understanding and challenges in this knowledge-stream to fully capitalize with its potentials. One of the contemporary challenges is to accurately verify data veracity, and developing value from the verified data for an organization and its stakeholders. Consequently, the purpose of this paper is to develop insights on how the authors could strategically deal with the contemporary challenges in strategic management of big data, related to data veracity and data value.

Design/methodology/approach – The inductive–constructivist approach is followed to develop insights at the intersection of dynamic capabilities theory and stakeholder relationship management concept, in order to strategically deal with the contemporary challenges in big data management, related to data veracity and data value.

Findings – At the intersection of dynamic capabilities theory and stakeholder relationship management concept, an implication is acknowledged, which has research and practical significance to strategically verify data source, its veracity and value. Following this implication, a framework of a data incubator is proposed to proactively develop insights on veracity and value of data. Empirical insights are also presented in this study to support this initial framework.

Practical implications – For future research in strategic management of big data, academics will have contextual understanding on the particular interconnected and interdependent attributes from dynamic capabilities and stakeholder relationship management research streams to further enhance the understanding in big data management. For practice, these insights will be useful for executives to focus on specific attributes of the proposed data incubator to confirm data veracity and develop insights on how to design, deliver and evaluate stakeholder value based on the verified data.

Originality/value – Following a synthesis at the intersection of dynamic capabilities theory and stakeholder relationship management concept, this study introduces a data incubator to meaningfully deal with the big data management challenges, related to veracity and value of data.

Keywords Dynamic capabilities, Big data, Resource-based view, Stakeholder relationship management, Data incubator

Paper type Conceptual paper

Introduction

Big data analytics and management has received enormous attention from academics and practitioners of a number of knowledge streams, in general, because of its practical cross-functional contribution, and further cross-disciplinary implications and potentials to contribute to diverse socio-economic and ecological issues. However, the discussion in this paper on the extant research and practice of big data management demonstrates that we have limited knowledge on to fully capitalize on the potentials of big data analytics. In particular, the extant literature demonstrates 5Vs in the definition of big data analytics, which are high volume of data, intense velocity in the appearance of large amount of data, great but confusing variety in the perspectives and potentials of data, data veracity and data value, e.g. how particular data set could offer specific value to a firm and its stakeholders. However, the extant literature in this field argue that from the different management perspectives, diverse
challenges appear in big data exploration/identification, visualization, storage, pattern mining and analytics, which adversely impact on the contemporary decision making process, while exploring multifaceted potential implications of big data (Hilbert, 2016). It gives us a research and managerial problem to develop insights on how we could simplify the big data management structure, in order to fully exploit the full potentials of big data analytics.

In this context, this study aims to develop insights on how we could strategically deal with the contemporary challenges in big data management, related to data veracity and to data value. In order to contribute to the current understanding on how we could verify the source of data and data veracity, and once the data veracity is confirmed, how we could develop insights to strategically capitalize data, this study focuses on the dynamic capabilities theory and the stakeholder relationship management concept. On the one hand, the purpose of relying on the dynamic capabilities theory is, in general, the overall management of data veracity and data value management, both, generally requires to be managed by organizational processes and different types of firm resources to pursue the firm’s path/strategic direction, and such organizational processes, resources and path dependences are the fundamental constituents to form organizational dynamic capabilities (Teece et al., 1997). On the other hand, in general, an organization attempts to collect and analyze data about its stakeholders. developing a strategic management perspective of big data based on analyzing the stakeholder relationships and interactions appear as instrumental to design and deliver stakeholder value (Campanella et al., 2016; Ferraris et al., 2016) from the verified data sets. This is why, the stakeholder relationship management concept is also analyzed in this study, alongside the dynamic capabilities theory.

At the intersection of dynamic capabilities theory and stakeholder relationship management concept, this study proposes a data incubator (i.e. the framework in Figure 1) to verify the data source, i.e. from which stakeholders a particular data set is collected, how and when the data were collected and to develop insights how stakeholder value can be designed and delivered through such verified data, focusing on the stakeholders’ motivational variables. Such stakeholders’ motivational variables are their needs, wants and expectations, related to the market offerings (e.g. products or services) of an organization. This study also develops some...
empirical insights, in support of the proposed framework (Figure 1) to verify data veracity and understand data value.

In this context, this study analyze the existing literature on the consistent areas of relevant knowledge streams to develop insights on the problem area, in order to better understand how we could strategically and more meaningfully verify data veracity to ensure data value. As a result, this study pursues an inductive–constructivist method, to corroborate findings that evolve from the analyzed literature (Eisenhardt, 1989; Yin, 1994; Smart et al., 2012; Osman et al., 2014; Shams and Kaufmann, 2016). Therefore, this paper presents the pertinent literature along the progress of arguments as an inductive analysis, to justify the findings, in relation to the aim of the study (Glaser and Strauss, 1967; Hallier and Forbes, 2004; Randall and Mello, 2012; Naidoo and Wu, 2014). In line of this plan, this paper, first provides an overview from the extant literature to articulate the need for new research in big data management, its critical challenges in strategic management, and the relevant research questions and research aim. Second, the significance of the dynamic capabilities theory and the stakeholder relationship management concept is discussed to justify its joint implications to design a data incubator, in order to verify data veracity and practical value of data. Third, the strategic management framework for big data (Figure 1) is depicted and discussed. Finally, the contribution of the study is justified in the conclusion section.

Need for new research in big data management

On the one hand, big data management has engendered enormous potential in exploring new management insights and phenomena, in order to proactively and more meaningfully encounter the contemporary management challenges. As a result, big data analytics and management has received great attention in recent years from researchers and practitioners in different public and private sectors, in order to contribute to diverse socio-economic and ecological issues (Hargreaves et al., 2018; Vassakis et al., 2018). On the other hand, a number of studies and industry reports in recent years argue that we have limited knowledge in big data management to fully exploit its potentials, in order to develop new insights and to contribute to diverse socio-economic and environmental issues (Chen and Zhang, 2014; Newman and Farrell, 2015; Bello-Orgaz et al., 2016; Dubey et al., 2016; Rumsfeld et al., 2016; Barth, 2017; McDonald, 2017; O’Grady, 2017; Bikakis et al., 2018; Couture, 2018; McNamee and Parakilas, 2018). In support of this view, Mikalef et al. (2017) posit that:

[...] big data and analytics are also challenging existing modes of business and well-established companies. Yet, there is limited understanding of how organizations need to change to embrace these technological innovations, and the business shifts they entail. Even more, the business value and strategic relevance of big data and analytics technologies still remain largely underexplored. (np)

As a consequence, from the management perspective of big data, diverse challenges appear in big data exploration/identification, visualization, storage, pattern mining and analytics, which adversely impact on the contemporary decision making process related to business, healthcare and other sectors as well (Hilbert, 2016). It gives us a research and managerial problem to develop insights on how we could simplify the big data management structure, in order to fully exploit the full potentials of big data analytics. Also, streamlining the big data management structure is instrumental to avoid any management myopia and undertake the right decision, at the right time, based on the right data set, and targeting the right stakeholders, with an aim to contribute to different social and industry issues.

A critical challenge in big data management

The concept of big data defined by Gartner (2012, p. 6) is based on 3Vs: “big data is high volume, high velocity and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization”
The processing of such high volume of data that derived from diverse organizational sources of data, randomly in a high velocity, with rich variety of information that would have multifarious managerial implications; however, processing such enormous flows of large amount of data is a complex procedure, in general. Two additional Vs are added in the definition of big data, in order to develop insights on the significance of available data that comes in high volume, high velocity with colossal variety in perspective and prospective implications of such large amount of data. These two additional Vs are veracity and value of data (Yin and Kaynak, 2015; Rumsfeld et al., 2016).

The underlying strategic management challenges, associated with the purpose of adding these two Vs in the definition of big data is to understand how authentic the information that we (the researchers and practitioners) retrieve from such large amount of data, and how we could analyze the data promptly, in order to understand the real value of particular data set, in order to develop new data-based insights to contribute to the needs, wants and expectations of different associated stakeholders and diverse socio-economic and ecological issues. As a result, this study focuses on these two strategic management challenges in big data analytics, in order to understand the veracity and value of high volume of data that appears in high velocity and in extreme variety.

Veracity and value in strategic management of big data, and the research questions and research aim

Veracity is synonymous to authenticity. Strategically, authenticity is defined as “socially constructed concept” (Cohen, 1988, p. 374), which replicates as “genuineness, not being false[…] and of having verified origin” (Gundlach and Neville 2012, p. 485). Market offerings are often promoted based on its genuineness as an indication of authenticity or veracity of the claims and contributions of the market offerings (e.g. products or services) to the stakeholders’ needs, wants and expectations (Chhabra, 2005). As a result, in order to understand the real value of data, developing insights on the veracity of data is essential. Upon verifying the veracity of data, the next key concern is to understand the real-life socio-economic and/or ecological value/implication of that authentic data set. In broader strategic management literature, value is defined as:

[… an anticipated outcome of any sort of planned and organized activity. The activity could be derived from monetary, psychic, or physical resources. The more the outcome meets initial anticipation, the more the possibility of win-win outcomes or value optimization for all involved stakeholders. (Shams, 2013, p. 244)

Centered on these two strategic contexts of veracity and value, the following two research questions (RQ) are derived, in order to encounter this critical challenge in big data management:

**RQ1.** How the veracity of data can be verified or the source of data can be verified, in order to avoid any management myopia in the future?

**RQ2.** How analyzing the planned and/or organized activities of the associated stakeholders would be instrumental to understand the value of data and information?

Focusing on these two research questions, which are derived from the extant literature, this study aims to develop insights on how we could strategically deal with the contemporary challenges in big data management, related to data veracity and data value.

Dynamic capabilities theory: the strategic management perspective of big data

Teece et al. (1997) coined the dynamic capabilities theory, centered on the competitive business environment, as an extension of the resource-based theory of strategic management. The dynamic capabilities theory is evolved, focusing on some interconnected management
philosophies (Gupta et al., 2018; Xu and Wang, 2018; Vezina et al., 2018; Lee, 2018; Oliva et al., 2018). The market powers/drivers are the first one, which is instrumental to envisage a firm’s strategic goal, in order to structure or restructure the firm’s managerial processes. Analyzing the firm’s entire business environment, the overall socio-economic environment where the firm operates, its industry, target markets and other stakeholders is instrumental to ascertain the firm’s market powers/drivers. Second, another interrelated management concept in the dynamic capabilities theory is the resource-based view (RBV), which is devoted to strategically source and allocate firm’s resources, with an aim to attain competitive advantage for the firm. The firm’s existing resources enable its managerial processes to plan and implement its strategies, in order to enrich the firm’s operational capabilities and efficacy (Daft, 1983), centered on the firm’s extant and/or envisioned competitive position in its industry. The key firm resources are typically attributed in three different categories. The first one is physical resources of the firm (Williamson, 1975). It includes the physical plant and equipment. The second one is the collective skills and proficiencies of the firm’s complementors, employees, managers and the board of directors, as its human resources (Becker, 1964). The organizational resources (Tomer, 1987) of the firm is the third kind of firm resource is attributed the firm’s premediated and casual decision making procedure.

The dynamic capabilities theory is derived based on a strategic management concept to allocate/reallocate such firm resources in a simulative measure, with an aim to attain and sustain the sources of the firm’s competitive advantage. This simulative measure in organizational dynamic capabilities relies on the three different issues of the firm, which are “processes, positions and paths.” The processes here are denoted as the overall organizational and managerial routines of the firm. The availability and accessibility of the firm’s three different types of resources are referred as the position of firm’s resources here, the path here is elucidated as the strategic direction that the firm has followed so far and/or could follow in the future, as the path dependencies of the firm. In brief, Teece et al. (1997) discussed a strategic organizational routine to coordinate, integrate, learn, configure/reconfigure and transform organizational processes, on the basis of strategically allocating and/or reallocating three different types of firm resources, in order to pursue the path/strategic direction of the firm, so that the firm can strategically organize its market powers/drivers to proactively deal with its competitive business environment. This organizational routine is recognized as the dynamic capabilities theory in strategic management literature, which is an extended approach of the RBV to generate business revenue at a planned amount, and to deal with the business competition (Mukherjee et al., 2013; Singh et al., 2013; Anderson et al., 2015, Gaur et al., 2014; Contractor et al., 2016; Nuruzzaman, Gaur, and Sambharya, 2018; Nuruzzaman, Singh, and Pattnaik, 2018; Kumar et al., 2018).

Since, the overall management of data veracity and data value management, both, generally requires to be managed by organizational processes (e.g. day-to-day routine update of data and information, and/or alignment of data with strategic goal), and different types of firm resources (e.g. physical resources, human resources and organizational resources), in order to pursue the firm’s path/strategic direction, focusing on the dynamic capabilities theory is appeared as instrumental to leverage the management of data veracity and data value. The next two sub-sections of this section attempt to develop insights on this theoretical proposition.

First- and second-order dynamic capabilities
Arndt et al. (2018) accentuates the importance of the continuous process of organizational knowledge management to learn and purify the specific organizational dynamic capabilities to proactively respond to the rapid changes in the competitive market environment. Focusing on this importance of continuous organizational learning process about the business environment, more recently, Schilke (2014) argues about a segmentation of the
Overall organizational dynamic capabilities. This segmentation of dynamic capabilities first discusses about the first-order dynamic capabilities, which are the regular day-to-day organizational routines that focus on the operational resource-base of a firm. Second, this segmentation of dynamic capabilities discusses about the second-order dynamic capabilities, which are comparatively more complex capabilities to influence the first-order dynamic capabilities, in an attempt to attune with the changes in the competitive business environment. Peteraf et al. (2013) argue about the dynamic bundle of resources and capabilities. These arguments of Peteraf et al. (2013) associate with a notion that the first-order and second-order dynamic capabilities could be structured to evaluate the second-order dynamic capabilities, as a consequence of the first-order dynamic capabilities, in order to recognize how accurately these two types (bundle) of capabilities can be twisted for finest organizational results. In relation to this proposition, an example is discussed in the next sub-section of this section, in order to discuss the significance of the first- and second-order dynamic capabilities (bundle of capabilities) to establish a stakeholder-centered data incubator that would have implications to develop insights on data veracity and data value.

First- and second-order dynamic capabilities (bundle of resources/capabilities), and stakeholder-centered data incubator

Since an organization attempts to collect and analyze data about its stakeholders, developing a strategic data incubator, on the basis of analyzing the “cause and consequence of stakeholder relationships and interactions as a stakeholder causal scope (SCS)” (Shams, 2016a, p. 141), would be instrumental to develop insights on data veracity and the value of collected data. For the purpose of this study, data incubator could be defined, in relation to the universal definition of incubator. For example, data incubator would be an analytical system or computer software that attempts to generate new insights from a large amount of complex, but premature volatile raw data. From this context, a stakeholder-centered data incubator would be useful to analyze data, in relation to diverse range of different SCSs among various stakeholders to intertwine the organizational knowledge management and learning efforts based on the first- and second-order dynamic capabilities (i.e. the bundle of capabilities) and in order to verify the source of data, and its veracity and value. For example, a more complex process, such as business partnership formation, merger and acquisition decision or product innovation initiative as the second-order dynamic capabilities can be envisioned and executed, in support of the first-order dynamic capabilities (e.g. day-to-day organizational routine), together as a dynamic bundle of first- and second-order dynamic capabilities/resources. In this intertwining sequence of the first- and second-order dynamic capabilities, the day-to-day organizational routine, usually could contribute to a firm’s organizational learning processes, centered on the regular operational experiences (e.g. service encounters, customers’ transactions etc.). Such operational experiences would be instrumental to reinforce the second-order dynamic capabilities (e.g. product or service innovation) to uphold the firm’s competitive position. As a consequence, an organized and chronological influence of the first- and second-order dynamic capabilities, as bundle of capabilities/resources, would be able to proactively analyze SCSs to understand which data are collected from which stakeholders and its veracity, and what would be the value of that particular data or data set to that specific stakeholders. Sequentially, the organizational resources can be allocated and/or reallocated, on the basis of the newly developed insights on stakeholder-centered data veracity and data value, in order to plan and implement more complex processes (e.g. product development as second-order capabilities) to adapt with the changes in competitive business environment for ensuring organizational performance (e.g. sustained competitive advantage). Figure 1 depicts these dynamics of stakeholder-centered data incubator and the dynamic bundle of capabilities to
verify the origin/source of data, and the value of data (i.e. how the data can be manipulated to generate value) for the associated stakeholders (e.g. customers, business partners, suppliers and others).

**A strategic management framework for big data: discussion and empirical insights**

Figure 1 attempts to clarify the impact of the bundle of capabilities (first- and second-order dynamic capabilities) on data incubator to verify the source of data and its veracity, and how such authentic data can be manipulated to generate value for the associated stakeholders, including customers. The arguments of Figure 1 start from the top left-hand side’s square-box and the left-hand side’s square-box at the bottom. In general, an organization can collect data through its first-order dynamic capabilities, e.g. service encounter, customer interaction and so forth. For example, while an Australian university’s international student administrator was discussing with a new international student in their university to know the reasons of the students to select Australia as his international education destination, the student answered that “one of the reasons influencing my decision to study in Australia is its proximity to Singapore” (Tan, 2014, np; as cited in Shams, 2016b, p. 685). Based on such service encounters as the first-order dynamic capabilities, the university is able to collect data from their customers (international students) that “proximity” of the internationals students’ host country to their home country is one of the decision making factors of their customers. Such first-order dynamic capabilities confirmed the source of data about the “proximity” issue, in order to verify that this particular international student as the sample from the total international student population has confirmed that “proximity” is one of the variables in international students’ decision making process.

In terms of thinking about promotional innovation (e.g. new insightful and influential ways of product or service promotion) as the more complex procedure of second-order dynamic capabilities, the university can further verify the veracity/authenticity of this confirmed source of data, i.e. the “proximity”, based on analyzing the extant research findings in international education management and marketing. For example, the previous academic and industry-based studies already confirm that the international students from the Asian countries, which are the largest markets of the global education industry, consider “proximity” as one of their decision making factors, while selecting their international education destination/host country among the alternative international education destinations (Phang, 2013; Singh et al., 2013; UNESCO, 2014; International Students Australia, 2015). As a result, both, the first- and second-order dynamic capabilities as the bundle of capabilities collectively can verify the source and veracity of data (i.e. “proximity” in this case).

Upon, confirming data veracity, the next key concern is analyzing data to understand and confirm how data can be manipulated to develop value for the associated stakeholders. In general, analyzing the needs, wants and expectations of the stakeholders and their SCSSs (i.e. the cause and consequence of stakeholder relationship and management) would be instrumental to develop and deliver practical value to the stakeholders based on the authentic data or data source. Furthermore, the needs, wants and expectations of stakeholders usually could be interrelated with the stakeholders’ cause and consequence of relationships and interactions. For example, in the empirical case example of this study, on the one hand, the reason (cause) of international students is to find an international education host country that closer (proximate) to their home country, in comparison to all other competitive host countries. On the other hand, the reason (cause) here for the university would be to understand their customers’ (international students’) motivation (needs, wants and expectations) to promote their international education based on the
motivational factors of their customers. In this context, the role of data incubator should be developing a synthesis on how analyzing their SCSs with their stakeholders/customers could nurture insights on how they could design, promote, deliver and evaluate value from their verified data and information to their customers based on the customers’ needs, wants and expectations, similar to the “proximity” case example' customers-need-based value generation while verifying the source of data.

Sequentially, upon verifying the data veracity and the value of data, based on the dynamics of Figure 1, the organizational resources can be allocated and/or reallocated, based on the newly developed insights (i.e. “proximity” in this case) on stakeholder-centered data veracity and data value, in order to plan and implement more complex processes (e.g. promotional innovation in this case) to adapt with the changes in competitive business environment for ensuring organizational performance (e.g. enriching competitive advantage, higher revenue, customer satisfaction and so forth).

Conclusion

The motivation of this study is to strategically deal with the two contemporary challenges in big data management, which are data veracity and data value. Since organizations generally collect data about their stakeholders to deliver superior value to the stakeholders, this study focused on SCS analysis concept to analyze the cause and consequence of stakeholder relationships and interactions in order to understand which stakeholder is the source of which data set, and to develop insights how analyzing stakeholders’ motivation (needs, wants and expectations) would be instrumental to design and deliver stakeholder value, on the basis of verified data, and as per the stakeholders’ motivational variable(s). In this context, this study analyzed SCS through the lens of dynamic capabilities theory. In particular, on the basis of the bundle of capabilities (synthesis of first- and second-order dynamic capabilities), this study proposes a stakeholder-centered data incubator (i.e. the framework in Figure 1) to contribute to our extant understanding on how we could proactively confirm data veracity and design stakeholder value from the verified data.

In support of the overall arguments of the study and the discussed dynamics of Figure 1 to confirm data veracity and data value, the study develops empirical insights, on the basis of the case example of “proximity,” which answers RQ1 and RQ2, and meets the research aim on strategically dealing with the contemporary challenges in big data management, related to data veracity and data value. The insights of this study offer implications for both research and practice in strategic management of big data. For future research in big data management, academics will have contextual understanding on the particular interconnected and interdependent attributes from dynamic capabilities and stakeholder relationship management research streams to further enhance our understanding on big data management. For practice, these insights will be useful for executives to focus on specific attributes of the proposed data incubator to confirm data veracity and develop insights on how to design, deliver and evaluate stakeholder value on the basis of the verified data.

References


Further reading


Corresponding author

S.M. Riad Shams can be contacted at: Riad.Shams@urfu.ru

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Challenges with big data analytics in service supply chains in the UAE

Mehmood Khan

College of Business, Abu Dhabi University, Abu Dhabi, United Arab Emirates

Abstract

Purpose – The purpose of this paper is to study the challenges associated with big data analytics (BDA) in service supply chains in the United Arab Emirates (UAE).

Design/methodology/approach – A comprehensive questionnaire has been developed based on semi-structured interviews with different administrators and IT experts. In the second phase, data (n = 164) are collected from procurement, operations, administration and customer service staff in the UAE. In the third phase, responses are examined using principal component analysis to identify eight major challenges for big data. A structural model is developed to examine the significance of these dimensions to the notion of big data challenges in supply chains.

Findings – The statistical model shows 66 percent variance of response to BDA, which is caused by technical, cultural, ethical, operational, tactical, procedural, functional and organizational challenges. These are positively correlated measurement challenges with BDA in service supply chains.

Research limitations/implications – Service supply chain professionals and stakeholders believe that catering to the challenges with BDA must be a multi-faceted approach and not limited to specific practices.

Practical implications – The challenges with BDA should be taken into planning and implementation from a holistic perspective. The framework in this paper can have both theoretical and practical implications.

Originality/value – The contribution of this paper is to advance the understanding of BDA in service sector by viewing it from the perspective of different stakeholders.

Keywords Service supply chains, UAE, Analytics, Big data, Challenges

1. Introduction

“Big Data” is the buzzword of the day. Data analytics is crucial for the success of any enterprise in this contemporary information age. Enterprises worldwide emphasize the importance of data collection and analysis to efficiently run their businesses. Big data provides organizations with insights into their business processes and enables them to promote their products better by employing analytical techniques such as data mining (Ngai and Chau, 2009). The amount of data generated from the beginning of time until the last decade is probably generated every couple of days today (Smolan and Erwitt, 2012). This trend, in all probability, is set to continue as the increase in social media applications and ease of access to data through mobile devices is exponentially increasing the amount of data that are captured every second (Korte et al., 2013). This massive amount of online data is affecting supply chains all over the world.

The volume of this information enables experts in mathematics, statistics, computer science and behavioral sciences to assess future outcomes with better accuracy (Davenport and Patil, 2012). However, it is not merely the volume but the variety, veracity, value and velocity of data that challenge supply chains nowadays in a completely new arena of competition where predictive data analytics is the key to success (Waller and Fawcett, 2013). Undoubtedly, the usability of the data is largely determined by its analytical potential (O’Reilly, 1982). Poor data analytics can have a direct impact on business decisions (Warth et al., 2011) and may translate into massive losses (Dey and Kumar, 2010). The current state of literature on big data in management disciplines can easily be classified into four broad streams, i.e. theoretical development; management transition; capabilities
and performance; and supply chain management (SCM) (LaValle et al., 2011; Chen et al., 2012; Wamba et al., 2015; Mayer-Schönberger, 2015; Ardito et al., 2018).

However, despite the widespread promotion of big data, recent industry reports indicate that supply chain managers are apprehensive to invest in big data analytics (BDA), mainly because of their past failures with business intelligence packages (Nair and Narayana, 2012). To help practitioners resolve this dilemma, it is essential to understand some of the differences between BDA and traditional business intelligence tools. Big data is an all-encompassing term for data sets that include web content, news feeds, social media postings, video clips and so forth. These data sets with their appreciable sizes and varying complexity levels are difficult to capture, process and manage in a timely fashion using on-hand data management tools and traditional data processing applications (Snijders et al., 2012). Therefore, BDA, per se, has been recognized as a radical departure from the traditional business intelligence tools (Gillon et al., 2014).

Big data has the potential to revolutionize supply chain dynamics. The exponential growth in the quantity and diversity of data has led to the creation of data sets larger than what is manageable by the conventional and hands-on management tools. To manage these new and potentially invaluable data sets, new methods of data science and new applications would be required in the form of predictive analytics. These new applications will transform the way the supply chain system is designed and managed, presenting a unique and significant challenge for logistics and SCM.

BDA in SCM has received increasing attention due to the significant challenges that can potentially result in inefficiencies and wastage in supply chains, such as delayed shipments, rising fuel costs, inconsistent suppliers and ever-increasing customer expectations, among others (Barnaghi et al., 2013). Companies expect to capitalize significantly on BDA in logistics and supply chain operations to improve visibility, flexibility and integration of global supply chains and logistics processes to effectively manage demand volatility and handle cost fluctuations (Genpact, 2014). This would help managers make strategic decisions on sourcing, supply chain network design and product design and development. In the operational planning phase, data analytics would be useful in demand planning, procurement, production, inventory and logistics.

Generally, big data needs to be identified by a number of fundamental characteristics (Daniel, 2015):

- volume – enormous bulk of data;
- velocity – daunting flow of information within a supply chain;
- veracity – noise, abnormality and uncertainty in data;
- variety – diversity of structured and unstructured data; and
- value – insights and benefits within a supply chain.

According to stakeholder theory, supply chains that pursue data analytics must ensure that stakeholders are committed to face inherent challenges. In service supply chains, customers’ satisfaction and responses are critical to business operations. Therefore, employing BDA should be a key objective and an operational imperative within service supply chains. The available literature has focused on opportunities and benefits of data analytics (Waller and Fawcett, 2013; Singh et al., 2017); however, there are few studies conducted on the challenges faced in implementing big data, especially within service supply chains. This dimension has largely been emphasized for considering analytical issues or conducting literature reviews, but rarely for studying technical ramifications of organizational decisions to implement big data. Thus, supply chain managers are required to consider the personal/organizational/technical range of challenges implementing big data throughout the life cycle of their service (Zhong et al., 2016).
This paper aims to expand upon the literature by exploring those challenges from the stakeholders’ perspectives. More specifically, based on the service supply chains in the United Arab Emirates (UAE), the paper presents an exploratory and confirmatory analysis of the challenges involved in employing BDA. This effort is in line with the country’s vision – Vision 2021 – to transform the economy into a diversified knowledge-based one to sustain long-term economic growth and generate job opportunities for the country’s young and growing population (Schilirò, 2015).

Thus, the purpose of this paper is threefold:

1. to explore the challenges in employing BDA in the UAE’s service supply chains;
2. to propose a comprehensive framework to address these challenges; and
3. to investigate and validate the relationship between the dominant factors by introducing a second-order confirmatory factor analysis (CFA) model.

The paper is organized as follows. A review of the relevant literature is presented in Section 2 to describe the concept of BDA, its theoretical background, key challenges and insights into service supply chains. Section 3 presents the methodology of the paper. The analysis and findings are discussed in Section 4. Section 5 concludes the paper.

2. Literature review

The concept of supply chain signifies the flow of information along with the flow of material and money (Souza, 2014). With recent technologies, supply chains can monitor the flow of information and are inclined toward collecting and analyzing a variety of data for efficient management (Chae and Olson, 2013). A typical supply chain may have to manage the inflow of more than 100 gigabytes of data every day. In fact, radio-frequency identification (RFID) tags play a significant role in tagging this massive bulk of available information (Tachizawa et al., 2015). Social media is another big source of big data for service supply chains nowadays (Fawcett and Waller, 2014). It is believed that the volume of digital data would reach 35 zettabytes by 2020. BDA has the potential to dictate digital manufacturing, mass customization and adaptive services (Tien, 2015). This timely adoption of BDA could greatly improve supply chains’ capabilities in a dynamic market environment (Meredith et al., 2012). However, to effectively integrate big data technology into SCM, organizational and behavioral issues related to adoption and practice have to be addressed.

2.1 Theoretical background

The available literature on data analytics mostly delves into the technical nature of the issues surrounding supply chains. However, some traditional theories that can be used to respond to the research questions relating to BDA are the stakeholder theory, resource-based view (RBV), transaction cost economics (TCE) and systems theory (Figure 1).

Stakeholder theory. This theory is a benchmark to highlight the difference between stakeholder management and stakeholder accountability (Khan et al., 2018). The literature on stakeholder theory can be classified into three perspectives, i.e., corporate, a stakeholder and a conceptual (Steurer, 2006). The corporate perspective emphasizes on the business–society interface; the stakeholder perspective focuses specifically on the stakeholders’ point of view; while the conceptual perspective approaches the theory from a particular concept, such as: the common good, human rights, environmental protection, and sustainable development. In the age of smart devices, supply chains are trying hard to be sustainable (with respect to managing pollution, congestion and costs) at every stage. Thus, the interaction between modern communication tools and sustainability targets poses a critical challenge for all the stakeholders (Gürsoy and Yücelen, 2017).
Stakeholder theory argues that in the context of big data, organizations have higher power, legitimacy, urgency and salience than that of individuals and societies (Tona et al., 2018). This challenge urges stakeholders to understand and capitalize on big data through nontraditional engagements.

Resource-based view. While examining the impact of investments in information technology, the RBV of a firm has been one of the most employed theoretical perspectives over the past three decades (Wade and Hulland, 2004). The main argument of RBV is that resources that are valuable, rare, in-imitable and non-substitutable are the building blocks of a competitive advantage (Bharadwaj, 2000). In terms of resources in information technology, they have been distinguished into tangible (infrastructure), human (human skills and knowledge) and intangible (culture and relationships) (Bharadwaj, 2000). This has enabled researchers and practitioners to identify the different types of IT resources their firms should aim to acquire and strengthen big data (Mikalef et al., 2016). Therefore, the capability to quickly turn raw data into useful knowledge offers supply chains with sustainable competitive advantage (Fawcett and Waller, 2011). However, this capability is highly dependent on the quality of data. Supply chain managers may have to decide when and how often they need to clean big data to maintain accuracy of their knowledge resources. This signifies the importance of data quality in enhancing supply chain performance through data analytics (Hazen et al., 2014).

Transaction cost economics theory. Transaction may be defined as a transfer of a good or service across a technologically separable interface (Amit and Zott, 2001). TCE is one of the most widely used theoretical lenses in business research (Macher and Richman, 2008). It focuses on the amount of effort and resources required by the stakeholders in a supply chain for an exchange while minimizing the total transaction cost (Hazen et al., 2014). Value can be derived from the attenuation of uncertainty, complexity, information asymmetry and bargaining conditions (Amit and Zott, 2001). Besides, reputation, trust and transactional experience can also increase the efficiency of a transaction (Mbaluka, 2013). Thus, the trust between stakeholders in a supply chain is often dependent on big data and other resources and an appreciation of long-term returns being greater than potential short-term gains (Williamson, 2008).

Systems theory. Systems theory is another commonly used notion in the SCM literature (Chicksand et al., 2012) that provides a useful lens to view big data challenges. The quality of
data deteriorates as a system absorbs bulk of data at all nodes. Systems theory claims that
the internal structures of big data need to be observed in order to understand their effect on
any system. Systems theory also contributes to the idea that big data in their vastness
cannot be understood, influenced or even changed by any system at an organizational level.
Only the relevant big data in an organization can be affected and interfered with. Interfering, regulating and harnessing the relevant portions of big data would be of inherent
interest for an organization (Scholz, 2017). Therefore, the control of data quality is of prime
importance from a systems theory perspective when investigating the impact of BDA on
SCM performance.

In summary, this paper employs the above theories to explore the following challenges
in BDA:

- using big data for enhancing the mutual engagement between stakeholders in service
  supply chains – stakeholder theory;
- Using big data as a unique and rare resource that is irreplaceable in service supply
  chains – RBV;
- Using big data to minimize their internal and external transaction cost in service
  supply chains – TCE; and
- Using big data to assess the impact of the external environment on logistics and
  supply chain operations – systems theory.

2.2 Big data analytics in service supply chains

The notion of services is in contrast with that of manufacturing in that it offers intangible
products; does not store inventory; does not require a physical site; and offers service only on
demand. BDA can play a critical role in defining and aligning operations in a service supply
chain at strategic and tactical levels (Wang et al., 2016). This allows managers to aim
for profitability with an improved flexibility. That is, these analytics tools provide a complete
overview of changing market conditions, upcoming risks and available capabilities. These
tools also help managers enhance performance through planning, procurement, production
and distribution. Table I presents the relevant literature in big data at various stages of a
supply chain.

Planning. This stage plays a vital role in setting the road map for future prospects of a
supply chain (Chen and Blue, 2010). Managers are required to move from an individual to a
collaborative perspective in decision-making (Frisk and Bannister, 2017). They have started
using analytic tools to study their budgets/expenses, supplier profiles and demand
projections to meet strategic objectives (Scott et al., 2018). This helps them to outline
appropriate contracting terms based on cost modeling and optimal risk assessment
(Jain et al., 2016). BDA helps businesses to adopt an industry’s best practices, performance
measures and to avoid disruptions (Rajesh and Malliga, 2013; Chai and Ngai, 2015). In addition, modern social networks manage certain aspects of the relationship between
vendors and suppliers (Souza, 2014). This ensures that vendors select only those suppliers
that have a proven track record of supplying goods/services on time with the required
quality. They often use analytical hierarchical processes to evaluate suppliers using a
complex set of dimensions (Rajesh and Malliga, 2013). A rigorous evaluation helps in
providing differentiated low-cost products/services (Srinivasan et al., 2012).

Procurement. BDA helps vendors to evaluate risks that must be avoided. These risks can
be identified on publicly available social media channels associated with sourcing markets
(Kabak and Burmaoğlu, 2013). Modern supply chains devise contingency plans to respond
to the possible risks (Mishra et al., 2013; Souza, 2014). Another trend is to develop
mathematical models and optimization approaches in supply chain relationships to handle disruptions (Khan, 2013). Analytics is a powerful platform for analyzing suppliers' performance for better sourcing (Oruezabala and Rico, 2012). This enables a quick evaluation of suppliers' quality, delivery time and reliability (Walker and Brammer, 2012; Yeniyurt et al., 2013).

Production. Analytics also helps modern vendors develop a bottom line for their production costs (Heo et al., 2012). They can have insights on production capacity levels to maximize their productivity (Noyes et al., 2014) and enhance their competitive advantage through agility (Dubey et al., 2018). It also helps manufacturers decide the right mix of multiple products and identify techniques to eliminate waste (Sharma and Agrawal, 2012).

### Table I.

Use of big data in supply chain management

<table>
<thead>
<tr>
<th>Article</th>
<th>Planning</th>
<th>Procurement</th>
<th>Production</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen and Blue (2010)</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Scott et al. (2018)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jain et al. (2016)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chai and Ngai (2015)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ekici (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rajesh and Malliga (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Srinivasan et al. (2012)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frisk and Bannister (2017)</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Kabak and Burmaoglu (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mishra et al. (2013)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoetmulder (2014)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khan (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oruezabala and Rico (2012)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walker and Brammer (2012)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeniyurt et al. (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heo et al. (2012)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noyes et al. (2014)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharma and Agrawal (2012)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li et al. (2013)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leung and Chen (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borade et al. (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borade et al. (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Downing et al. (2014)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang and Lei (2012)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guo and Li (2014)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fernandes et al. (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guerrero et al. (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Santoro et al. (2018)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dubey et al. (2018)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swaminathan (2012)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Najafi et al. (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drexl (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ozdamar and Demir (2012)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vidal et al. (2013)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lei et al. (2011)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minis and Tatarakis (2011)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferraris et al. (2018)</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Souza (2014)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>LaValle et al. (2011)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Chen et al. (2012)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mayer-Schonberger (2015)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Wamba et al. (2015)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ardito et al. (2018)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Challenges with big data analytics
Thus, they can devise several alternate production plans (Li et al., 2013). Furthermore, analytics provides them with insights on scheduling problems while modeling the sequence of production and rework operations (Leung and Chen, 2013). They can design an optimization system to handle the most complex retail, wholesale and multiple-channel inventory system (Souza, 2014; Santoro et al., 2018). In case of consignment stocks and vendor-managed inventory systems, analytics supports them in making informed decisions. It also aids their response to varying demands (Downing et al., 2014). BDA is also used to address problems within a multi-tier distribution network whereas it determines optimal inventory levels while considering demand variability, lead time delays and service levels (Wang and Lei, 2012; Guo and Li, 2014). Therefore, it would be possible for managers to have a holistic view of inventory levels before deciding on an optimal level of safety stocks (Fernandes et al., 2013; Guerrero et al., 2013).

*Distribution.* Global transportation and logistics generate a massive amount of data as several stakeholders manage the distribution operations. The data generated by RFID tags, mobile devices and electronic data interchange transactions can be utilized for improving distribution operations (Swaminathan, 2012; Ferraris et al., 2018). As the data are generated from various sources (e.g., costs, forecasts, capacities), it is essential to use analytics tools to give flexibility to the design of a supply chain (Najafi et al., 2013). It is vital in distribution planning to optimize the routes as well as the sequence of nodes for both crew and equipment (Özdamar and Demir, 2012; Lei et al., 2011). Thus, an optimal route and sequence of nodes would consider traffic volume, delivery windows and tolls on routes (Vidal et al., 2013; Souza, 2014).

### 2.3 Challenges with big data analytics

Supply chains are experiencing an increasingly complex and competitive environment. There is an ever-increasing pressure to respond to national and global changes (Daniel, 2015). The decision-making required to respond to these rapid changes is complex and difficult at times, especially if the available resources are not sufficient to utilize an abundant set of information (big data).

*Technical challenges.* With competition intensifying between modern supply chains, collecting and integrating real-time data can be challenging at times. The first challenge is to identify and establish trusted sources of data (Behera et al., 2015). Once that stage is over, a system would be required to clean and fragment the data into smaller segments. Most businesses design their systems to prompt an early warning at this stage if it detects any defects, exceptions or abnormalities. Inability to catch such defects in time may result in failures or damages. Another challenge with the analytics in big data is to devise machines and algorithms that are fault tolerant. That is, to ensure the probability of failure is at an “acceptable” level (Madden, 2012). Nowadays, supply chains often operate with cloud computing, which aggregates multiple disparate sources with very large clusters of data. This requires cloud storage to be scalable to execute various jobs to meet the goal of each cluster effectively (Katal et al., 2013).

As the bulk of information flows between stakeholders, uncertainty, errors and missing values are inevitable and must be managed. While sales and marketing heavily rely on crowdsourcing, the participatory sensors record data about location, speed, time, direction and even audios and videos. This challenges analysts with errors resulting from data inconsistency and incompleteness.

Another big challenge that data analysts come across is the heterogeneity of the unstructured data. This is so important that a processing error at one step can render subsequent analyses useless. Therefore, they require their systems to carry the provenance of data and its metadata through the data analysis pipelines (Kache and Seuring, 2017).
**Cultural challenges.** Currently, social media and digitized archives have produced a massive quantity of data; however, analysts still face challenges in extracting the actual meaning of the data originating from different cultures (Bail, 2014). Many archived texts are conversations between individuals, groups or even organizations instead of responses to direct questions from data analysts who have a post hoc intuition about the underlying factors. A higher level of quality can be achieved in analytics if executive sponsorship is assured from the beginning (Barlow, 2013). In other words, coordination between the business, technology and analytics executives is crucial to achieve desired results. The whole paradigm of big data is in its nascent stage and it may take a while for managers to have that level of trust in big data to subscribe to the idea (Kache and Seuring, 2017). Another agonizing issue for business executives is the ownership of the data, while there are petabytes of data available on social networking sites, neither the sites nor the users fully own the data. This dichotomy becomes even more complicated if an organization is depending on cloud computing (Kaisler et al., 2013).

**Ethical challenges.** With an abundance of online social networks, privacy of data is becoming a pressing concern for modern supply chains. There is a general public fear regarding the misuse of personal data, particularly when data are linked from multiple sources (Jagadish et al., 2014). This challenge must be addressed in the context of customer satisfaction. Nowadays, service providers face a formidable challenge in differentiating between commercial and public uses of big data while keeping individual rights and liberties of their customers in mind (Vayena et al., 2013). Service supply chains can reduce their role in surveillance by becoming more visible to consumers and by limiting their data collection. This policy would also minimize the roles of the firms that are invisible to consumers, i.e., data aggregators and data brokers (Martin, 2015). Particularly, in the context of data derived from social network interactions, the consumers are mostly unaware of the ways their data can be used. Thus, to navigate this uncertain environment, diligence must be exercised about data provenance and there should be transparency maintained about their ultimate use. One may also argue the consumers’ right to question the very process of collecting their private data. For the sake of ensuring legitimacy, supply chains must oblige individual rights and dignity. In other words, the pursuit of big data should strictly follow a standard to ensure that risks and costs to individual consumers are proportional to benefits (Vayena and Tasioulas, 2013).

**Operational challenges.** The concept of big data can be implemented successfully in service supply chains only when stakeholders collaborate. This is consistent with the view that the value of big data will be based on the ability to co-create governing structures and ownership of the challenge involving supply chain performance. However, there would always be a divide in the opinions of supply chain executives regarding the scope of data extraction and utilization. The real benefits of this initiative can only be reaped if the participants are flexible enough to make their analytics more accessible (Dyckhoff et al., 2012). Nevertheless, the risk of possible misuse of data will always be present. As the quantity of data is ever growing, the participants require a vigorous learning management system to ensure reliable warehousing, transparent extraction and responsible reporting (Dringus, 2012).

**Tactical challenges.** The whole spectrum of big data is so dynamic that it has undermined the standard premise of the strategy adopted for long-term commitments. Although service supply chains have started utilizing consumer profiling methodologies in their marketing strategies, outlining a segmentation strategy for a longer period is yet another complex task that depends on the clusters’ internal and external data (Bhimani and Willcocks, 2014). The industry, generally not used to receiving such bulk of information, continuously struggles to make policy interventions (Aaltonen and Tempini, 2014). Furthermore, modern
supply chains would always be threatened by the entry of new competitors in the market that may have a daunting effect of their routine tactics and long-term strategies (Kache and Seuring, 2017).

**Procedural challenges.** Although modern systems utilize advanced encryption algorithms for keeping data confidential, attacks can easily destroy, expose, modify, disable or steal data by exploiting its vulnerability (Sivarajah et al., 2017). Another issue with the current information management systems is the speed of transmission that depends heavily on the bandwidth of a channel and protocol that defines the data structure over the channel (Daniel, 2015). Thus, a supply network would always have to keep their network protocols up to date to avoid hacking. A recent trend in the industry has been to use biological mechanisms such as DNA, which is unique and difficult to decrypt without an authorized key to adopt new protocols as well as to transmit data at a high speed. However, the astounding growth and multiplicity of unstructured data have intensely affected the way managers interpret new knowledge from the raw data. Thus, the middle managers as well as top executives will be required to keep themselves updated with the evolving analytical skills to interpret the current state-of-the-art data solutions (Phillips-Wren and Hoskisson, 2015).

**Functional challenges.** There is an enormous cost associated with developing algorithms to integrate structured and unstructured formats of data coming from distinct sources and stored in different departments of the stakeholders. Cleaning such massive quantities may result in the loss of data (Daniel, 2015). At times, suppliers and distributors do not have interoperable systems. Therefore, aggregating their administrative data also poses an additional challenge (Wagner and Ice, 2012). Furthermore, modeling such a disorderly, interrelated, but untrustworthy data is a complicated job as it is expected to unveil more reliable hidden patterns and knowledge (Jagadish et al., 2014).

**Organizational challenges.** There is a big range of challenges associated with making users understand that big data calls for new processes and new changes. These changes require extra budgets for collecting, storing, and mining large data sets (Daniel, 2015). The value of Big Data is also linked to the governing structures prevailing with the different stakeholders in a supply chain (Daniel and Butson, 2013). In fact, the stakeholders would have to use a myriad of electronic devices to store the ever-growing data so that queries, analysis and visualizations of the patterns can be executed (Zhong et al., 2016).

### 3. Research methodology

The focus of this paper is to explore the challenges faced in employing BDA and propose a framework to address these challenges in service supply chains. To this end, the paper has been divided into three phases. In the first phase, the paper has adopted an exploratory approach to collect empirical data from service organizations in the UAE. In total, 25 administrators and/or experts from various service organizations in the different emirates of the UAE were interviewed to explore the challenges in their organizations. To cater to an entire supply chain in services, experts were chosen from procurement, operations, administration and customer service. Their responses were filtered into a compiled list of 30 challenges they faced with analytics in their organizations (Table II).

The second phase adopted a quantitative methodology to generate an instrument (scale) that measures the constructs of the challenges of BDA. This is an important step given that the literature is fragmented with regards to challenges from a supply chain perspective. An instrument with 30 questions with a five-point Likert scale was used for conducting a pilot study. Conducting a pilot study provided the preliminary information about the reliability and validity of the measurement scales. In total, 400 individuals were invited to respond to the instrument through SurveyMonkey, and 164 complete responses (41 percent) were
received from procurement, operations, administration and customer service staff in
supply chains across the UAE. An overview of the demographics of this sample is outlined
in Table III.

The third phase was to test these responses. Principal component analysis (PCA) was
used to explain the maximum amount of common variance with the smallest number of
explanatory constructs (factors or latent variables). These factors represent clusters of
challenges that correlate highly with each other. A Cronbach’s $\alpha$ value ranged from 0.683 to
0.869 showed that the responses are reliable enough for further analysis. CFA was used to
assess and validate the constructs that describe big data challenges in supply chains. This
three-phase methodology is depicted in Figure 2.

4. Data analysis
The third phase was to test these responses. PCA was used to explain the maximum amount
of common variance with the smallest number of explanatory constructs (factors or latent
variables). These factors represent clusters of the challenges that correlate highly with each other. A Cronbach’s $\alpha$ value of 0.868 showed that the responses are reliable enough for further analysis. CFA was used to assess and validate the constructs that describe challenges in the service organizations.

4.1 Exploratory factor analysis (EFA)
This analysis was used in the pilot study, which involved doctors, administrators and other
stakeholders in the service organizations. Several conditions had to be met prior to testing

<table>
<thead>
<tr>
<th>Challenges with big data analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical challenges</td>
</tr>
<tr>
<td>1. Real-time data acquisition</td>
</tr>
<tr>
<td>2. Real-time processing and visualization</td>
</tr>
<tr>
<td>3. Fault tolerance</td>
</tr>
<tr>
<td>4. Scalability</td>
</tr>
<tr>
<td>5. Inconsistency and incompleteness</td>
</tr>
<tr>
<td>6. Heterogeneity</td>
</tr>
<tr>
<td>Cultural challenges</td>
</tr>
<tr>
<td>7. Lack of awareness</td>
</tr>
<tr>
<td>8. Lack of executive support</td>
</tr>
<tr>
<td>10. Lack of ownership</td>
</tr>
<tr>
<td>Ethical challenges</td>
</tr>
<tr>
<td>11. Privacy and security</td>
</tr>
<tr>
<td>12. Differentiation of commercial and public use</td>
</tr>
<tr>
<td>14. Transparency</td>
</tr>
<tr>
<td>15. Legitimacy</td>
</tr>
<tr>
<td>Operational challenges</td>
</tr>
<tr>
<td>17. Accessibility of analytics</td>
</tr>
<tr>
<td>18. Learning management system</td>
</tr>
<tr>
<td>Tactical challenges</td>
</tr>
<tr>
<td>Procedural challenges</td>
</tr>
<tr>
<td>22. Vulnerability of data</td>
</tr>
<tr>
<td>Functional challenges</td>
</tr>
<tr>
<td>27. Data modeling</td>
</tr>
<tr>
<td>Organizational challenges</td>
</tr>
<tr>
<td>29. Insufficient storage</td>
</tr>
<tr>
<td>30. Governing structures</td>
</tr>
</tbody>
</table>

Table II. List of challenges with big data analytics
whether the items were suitable to run the analysis. The tests included the Kaiser–Meyer–Olkin (KMO) test and Bartlett’s test of sphericity. The results are shown in Table IV.

The KMO test (KMO > 0.50) showed that these items are suitable for factor analysis and that there is no problem of serious multicollinearity in the data. Bartlett’s test of sphericity (sig. < 0.05) showed that the correlation between items is sufficient to run factor analysis. After the two tests, EFA was used with a PCA extraction method and varimax rotation on the 30-item instrument. The number of factors to retain was based on a combination of methods (e.g. eigenvalue > 1.0, scree plot) and the theoretical basis of the rotated factors. Items should preferably load greater than 0.30 on the relevant factor and less than 0.40 on all of the other factors (Stevens, 1992).

The next step was to analyze the content of the questions that load highly on the same factor to try to identify common themes (Table V). This factor pattern and their item loadings are presented in a rotating matrix component matrix in Table VI. The next step is to look at the content of questions that load highly on the same factor to try to identify common themes. The questions that load highly on factor 1 seem to relate to consistency, scalability and visualization and therefore were labeled as technical challenges. This factor consists of six items and accounts for 13.15 percent of the variance. The questions that load highly on factor 2 seem to relate to trust and support and, therefore, were labeled as cultural challenges. This factor consists of four items and accounts for 8.39 percent of the variance. The questions that load highly on factor 3 all seem to relate to privacy and transparency and were labeled as ethical challenges. This factor consists of five items and accounts for 8.27 percent of the variance. Similarly, the questions that load highly on Factor 4 seem collaboration and accessibility in stakeholders and, therefore, were labeled as operational challenges. This factor consists of three items and accounts for 7.89 percent of the variance. Following the same notion, the questions that load highly on factor 5 seem to relate to policy and competitors; and therefore was labeled as tactical challenges. This factor consists of three items and accounts for 7.66 percent of the variance. The questions that load highly on factor 6 seem to relate to speed and skills; and were labeled as procedural challenges. This factor consists of three items and accounts for 7.09 percent of the variance.
Challenges with big data analytics

Figure 2. Design/methodology of research

Table IV. KMO test and Bartlett’s test

| Kaiser–Meyer–Olkin measure of sampling | 0.748 |
| Adequacy Bartlett’s test of sphericity | $\chi^2$ |
| | 2,200 |
| Df | 435 |
| Sig | 0.000 |

Table V. Exploratory factor analysis

<table>
<thead>
<tr>
<th>Type of challenge</th>
<th>Number of items</th>
<th>Extracted variance (%)</th>
<th>Cumulative variance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical</td>
<td>6</td>
<td>13.15</td>
<td>13.15</td>
</tr>
<tr>
<td>Cultural</td>
<td>4</td>
<td>8.39</td>
<td>21.55</td>
</tr>
<tr>
<td>Ethical</td>
<td>5</td>
<td>8.27</td>
<td>29.81</td>
</tr>
<tr>
<td>Operational</td>
<td>3</td>
<td>7.89</td>
<td>37.70</td>
</tr>
<tr>
<td>Tactical</td>
<td>3</td>
<td>7.66</td>
<td>45.36</td>
</tr>
<tr>
<td>Procedural</td>
<td>3</td>
<td>7.09</td>
<td>52.46</td>
</tr>
<tr>
<td>Functional</td>
<td>3</td>
<td>6.55</td>
<td>59.01</td>
</tr>
<tr>
<td>Organizational</td>
<td>3</td>
<td>6.51</td>
<td>65.53</td>
</tr>
<tr>
<td>Items</td>
<td>Technical challenges</td>
<td>Cultural challenges</td>
<td>Ethical challenges</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Q23 Real-time data acquisition</td>
<td>0.867</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q22 Real-time processing and visualization</td>
<td>0.851</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q24 Fault tolerance</td>
<td>0.800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q13 Scalability</td>
<td>0.798</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q14 Inconsistency and incompleteness</td>
<td>0.749</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q15 Heterogeneity</td>
<td>0.395</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8 Lack of awareness</td>
<td>0.787</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9 Lack of executive support</td>
<td>0.756</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7 Lack of trust</td>
<td>0.682</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2 Lack of ownership</td>
<td>0.523</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q20 Privacy and security</td>
<td>0.636</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q21 Differentiation of commercial and public use</td>
<td>0.628</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1 Limited surveillance</td>
<td>0.618</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3 Transparency</td>
<td>0.589</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q19 Legitimacy</td>
<td>0.521</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8 Collaboration between stakeholders</td>
<td>0.807</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5 Accessibility of analytics</td>
<td>0.791</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4 Learning management system</td>
<td>0.730</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q30 Segmentation strategy</td>
<td>0.851</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q29 Policy interventions</td>
<td>0.781</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q28 New competitors</td>
<td>0.682</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q18 Vulnerability of data</td>
<td>0.755</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q17 Speed of transmission</td>
<td>0.707</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q16 Lack of analytical skills</td>
<td>0.629</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q26 Data cleansing</td>
<td>0.767</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q25 Data integration</td>
<td>0.755</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q27 Data modeling</td>
<td>0.480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q11 Insufficient budget</td>
<td>0.789</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q10 Insufficient storage</td>
<td>0.637</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q12 Governing structures</td>
<td>0.615</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average variance extracted (%)</td>
<td>13.1</td>
<td>8.4</td>
<td>8.3</td>
</tr>
</tbody>
</table>
Similarly, the questions that load highly on factor 7 seem modeling and integration and, therefore, were labeled as functional challenges. This factor consists of three items and accounts for 6.55 percent of the variance. Finally, the three questions that load highly on Factor 8 contain some components of governance and budgets and, therefore, were labeled as organizational challenges. This factor accounts for 6.51 percent of the variance.

4.2 Internal consistency and content validity

The degree of consistency of the responses over a construct is referred to as its reliability. The reliability coefficient, Cronbach’s $\alpha$, is generally used for this test. As shown in Table VII, the Cronbach’s $\alpha$ for the eight latent constructs indicates that the suggested constructs exhibit good psychometric properties.

**Convergent validity.** Convergent validity can be evaluated by the use of the Bentler–Bonett’s normed fit index (NFI). This index provides the degree to which the different approaches to measure a construct generate the same results (Ahire et al., 1996). According to a generally accepted principle, the NFI values of 0.90 or above are considered to be a satisfactory fit index (Bentler, 1992). As is shown in Table VII, the items in each construct converge well for further analysis.

**Discriminant validity.** Discriminant validity is the degree to which different latent constructs and their indicators can be distinguished from the other constructs and their indicators (Bagozzi et al., 1991). To calculate the discriminant validity, the Cronbach’s $\alpha$ of a latent construct is compared with its mean correlations with other latent constructs. A significant difference between these two measures is an indicator of discriminant validity (Ghiselli et al., 1981). As is shown by the values in Table VII, the five constructs are conceptually distinct.

4.3 Confirmatory factor analysis (CFA)

The exploratory factor analysis in this research identified eight challenges with BDA: technical, cultural, ethical, operational, tactical, procedural, functional, and organizational as a priori factors in service supply chains. Figure 3 shows a structural model of these factors causing an impact on the endogenous factor “big data challenges.”

A test of this structural model hints an acceptable goodness of fit ($\chi^2$/df = 718.4; CFI = 0.831; TLI = 0.814 and RMSEA = 0.071). Furthermore, as shown in this second-order model, all of the coefficient estimates of technical ($\beta = 0.29$; $p < 0.05$), cultural ($\beta = 0.66$; $p < 0.05$), ethical ($\beta = 0.76$; $p < 0.05$), operational ($\beta = 0.60$; $p < 0.05$), tactical ($\beta = 0.50$; $p < 0.05$), procedural ($\beta = 0.66$; $p < 0.05$), functional ($\beta = 0.71$; $p < 0.05$) and organizational ($\beta = 0.66$; $p < 0.05$) challenges, which describe the relationships or paths of the eight factors on the higher-order construct of big data challenges, are significant. It can be seen that the highest and the lowest impact appear to be from ethical challenges and technical challenges, respectively.

<table>
<thead>
<tr>
<th>Factors/challenges</th>
<th>Reliability Cronbach’s $\alpha$</th>
<th>Convergent validity Bentler–Bonett NFI</th>
<th>Discriminant validity factor Cronbach’s $\alpha$ – avg. correlation between factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical</td>
<td>0.869</td>
<td>0.969</td>
<td>0.684</td>
</tr>
<tr>
<td>Cultural</td>
<td>0.760</td>
<td>0.999</td>
<td>0.575</td>
</tr>
<tr>
<td>Ethical</td>
<td>0.683</td>
<td>0.899</td>
<td>0.498</td>
</tr>
<tr>
<td>Operational</td>
<td>0.776</td>
<td>1.000</td>
<td>0.591</td>
</tr>
<tr>
<td>Tactical</td>
<td>0.774</td>
<td>1.000</td>
<td>0.589</td>
</tr>
<tr>
<td>Procedural</td>
<td>0.693</td>
<td>1.000</td>
<td>0.508</td>
</tr>
<tr>
<td>Functional</td>
<td>0.695</td>
<td>1.000</td>
<td>0.510</td>
</tr>
<tr>
<td>Organizational</td>
<td>0.718</td>
<td>1.000</td>
<td>0.533</td>
</tr>
</tbody>
</table>

*Table VII. Construct validity analysis*
5. Discussion and implications
This paper aims at highlighting several inherent practices that challenge the implementation of BDA in service supply chains. Although the contemporary trend entices businesses with its charms, several organizations struggle with its implementation. An important aspect in this struggle is to bring all the involved stakeholders on board.
The literature does not offer a piece of research to light this dim but lucrative journey. Therefore, BDA is prevented from being effectively undertaken in practice and/or research.

This paper brings empirical evidence on how UAE-based service supply chains perceive big data challenges. An attempt has been made to identify the factors and quantify their impact on a service supply chain’s decision to integrate big data to their business. To this end, this paper is the first to propose a framework based on the RBV, TCE, stakeholder theory and systems theory. In light of these theories, the paper examines the direct impact of eight major challenges on the implementation of BDA in service supply chains. This provides new insights into how stakeholders and their interactions drive big data and institutionalize it into the fabric of a service supply chain.

The key research objectives were: to explore the challenges in the UAE’s service supply chains; to propose a comprehensive framework of challenges in service supply chains; and, to investigate the impact of these factors on big data challenges by using a structural (second-order CFA) model. The criterion-related validity of the measures on the framework was investigated by using convergent and discriminant validity of the instrument. This leads to a high level of confidence on the new scale.

Factor analysis resulted in an eight-factor model of measures with a total of 30 items. This analysis helped in the development of a framework that includes technical, cultural, ethical, operational, tactical, procedural, functional and organizational challenges. As noticed in Figure 3, all the eight factors have a significant impact on big data challenges. It is to be noted that the highest and lowest impacts come from ethical and technical factors, respectively. It explains the evolution of supply chains in response to the dramatic changes coming along with technology. Furthermore, this suggests a need for greater awareness among stakeholders to motivate people and groups to engage in the whole process of adopting big data development. However, despite a comparatively lower impact, technical factors still play a critical role in causing challenges for big data.

The paper confirms that it is vital for today’s service supply chains to have a comprehensive and holistic view while adopting big data. It would be necessary to weigh the significance of several factors and items identified in this paper in the context of various industries. The real fruits of Big Data would not be reaped without a support from every stakeholder. The findings reveal that organizations must follow a comprehensive and coherent approach to develop more supportive perceptions toward big data in service.

5.1 Theoretical implications
The emerging science of big data has so much in line with the current state of research and practice in information systems. This contextual support helps researchers and practitioners in suggesting a conceptual framework, testing a theory, forecasting a behavior and shedding light on an emerging theory (Rai, 2016). This paper provides a holistic view of the challenges with big data in supply chains. This holistic view helps researchers in:

- detecting irregularities across several contexts of big data in supply chains;
- developing a similar framework with one or all of the theories elaborated in the paper;
- analyzing several streams of unstructured data qualitatively to generate a new theory; and
- advancing the research in this emerging stream by highlighting the role of several challenges in one framework.

5.2 Practical implications
Despite a growing focus on big data in modern supply chains, managers feel a thirst to hit the underpinning challenges that the disarrayed bulk of information simply pours upon
their daily routines. This paper brings along a wide range of challenges by pointing out literally almost every possible hurdle that is faced by an operations or supply chain manager today. The framework in the paper helps them in:

- building and streamlining their data infrastructures from several stakeholders;
- differentiating between data analytics and traditional business intelligence tools;
- adopting newer applications in data science in the form of predictive analytics;
- using the bulk of information to boost visibility, flexibility, and integration;
- managing demand volatility and cost fluctuations proactively; and
- making strategic decisions while sourcing their raw material and designing their products.

5.3 Limitations and future research directions
Though the framework in this paper is linked with the challenges in service industries, it could very well be suitable for use in many other industries. Evidently, this framework is a valid and effective instrument to measure the challenges that will guide the efforts of practitioners to integrate big data. The broader scope of the study will help the UAE’s service practitioners to lead the transition toward a comprehensive approach to support the technological transition. The service facilities in the UAE must take part in the nationwide movement toward a more tech-friendly future. The paper has some limitations that may entice researchers for future studies. For example:

- investigating the contrast the challenges in manufacturing organizations;
- comparing the similar challenges in the neighboring countries to widen the scope of this study; and
- developing a numerical index of the challenges for a firm/industry with the help of the items highlighted in Table III.

References
Barlow, M. (2013), The Culture of Big Data, O’Reilly Media, CA.


Challenges with big data analytics


Further reading


Corresponding author
Mehmood Khan can be contacted at: mehmood.khan@adu.ac.ae
How do different types of interorganizational ties matter in technological exploration?

Yu-Shan Su  
Department of Industrial Education,  
National Taiwan Normal University, Taipei, Taiwan, and

Wim Vanhaverbeke  
Department of Business Economics,  
Universiteit Hasselt – Campus Diepenbeek, Diepenbeek, Belgium

Abstract

Purpose – Boundary-spanning exploration through establishing alliances is an effective strategy to explore technologies beyond local search in innovating firms. The purpose of this paper is to argue that it is useful to make a distinction in boundary-spanning exploration between what a firm learns from its alliance partners (explorative learning from partners (ELP)) and what it learns from other organisations (explorative learning from non-partners (ELN)).

Design/methodology/approach – The authors contend that alliances play a role in both types of exploration. More specifically, the authors discern three types of alliances (inside ties, clique-spanning ties and outside ties) based on their role vis-à-vis existing alliance cliques. Clique members are highly embedded, and breaking out of the cliques through clique-spanning and outside alliances is crucial to improving explorative learning. Thereafter, the authors claim that clique-spanning ties and outside ties have a different effect on ELN and ELP.

Findings – The empirical analysis of the “application specific integrated circuits” industry indicates that inside ties have negligible effects on both types of explorative learning. Clique-spanning ties have a positive effect on ELP, but not on ELN. The reverse is true for outside ties. The results show that research on explorative learning should devote greater attention to the various roles alliance partners and types of alliances play in advancing technological exploration.

Originality/value – The literature only emphasises the learning from partners, focussing mainly on accessing their technology. In sum, alliance partners play different roles in exploration, and their network position influences the role they are able to play.

Keywords Explorative learning from partners, Explorative learning from non-partners, Boundary-spanning exploration, Inside ties, Clique-spanning ties, Outside ties, Conduits, Prisms

Paper type Research paper

1. Introduction

Firms rely increasingly on external partners to explore new technologies (e.g. Rosenkopf and Nerkar, 2001; Lavie and Rosenkopf, 2006; Eriksson et al., 2016; Lavie et al., 2011; Stettner and Lavie, 2013). Explorative learning implies that firms assimilate and integrate knowledge that resides outside its corporate boundaries. Furthermore, there is convincing empirical evidence that technology alliances are instrumental for technological exploration (Gulati, 1998; Lavie and Rosenkopf, 2006; Lin et al., 2007).

Innovating firms can explore external technologies in different ways. The literature has emphasised the role of explorative learning through alliances and other formal interorganizational relationships. However, firms can also explore new technologies through their exposure to scientific publications (McMillan et al., 2000), patent releases, contact with consultants, technology providers and intermediaries (Howells, 2006), product introductions in the market, conferences, exhibitions, benchmarking with competitors (Wiersema and Bowen, 2008; Hunt and Morgan, 1997), mobility of personnel, etc. These are just a few examples to illustrate how firms explore
new technologies by relying on knowledge from organisations with which they have no formal innovation ties.

Given these multiple external sources of explorative learning, we suggest refining the concept of boundary-spanning explorative learning, as it was first introduced by Rosenkopf and Nerkar (2001), in two consecutive steps. First, boundary-spanning explorative learning can be split into two categories with respect to a firm’s existing formal relationship network: in boundary-spanning explorative learning, innovating companies learn from their partners (Schotter et al., 2017). However, they can also learn from organisations with which they have not established any formal relationships. Second, this distinction allows us to define two roles of alliance partners in firms’ explorative learning. First, a firm can learn from its alliance partners’ technology (we label this explorative learning from partners or ELN). Learning from external partners is not new and has been discussed extensively in the literature. Second, a firm can also explore new technologies based on knowledge from firms with which it has not established any formal relationship in the past (we label this explorative learning from non-partners or ELN). In ELN, technology-sourcing partnerships may foster exploratory learning by bringing a firm into contact with interesting sources of technology. Partners are thus acting as a reputation reference, or by informing a focal firm about new technological opportunities beyond its partners’ network (Podolny, 2001; Gulati, 1998; Ghosh and Rosenkopf, 2014; Singh et al., 2015). To date, innovation management literature has emphasised the role of alliance partners as external sources of knowledge (or ELN) and has neglected ELN.

More specifically, we are interested in finding out what type of formal agreements is conducive to both types of explorative learning. Two key questions have to be addressed:

RQ1. How a firm’s formal interorganizational relations do contribute to the two types of explorative learning?

RQ2. Do different types of formal relations play a different role in both types of explorative learning?

To provide an answer to the second question, we classify technology partnerships into three groups – inside ties, cross-spanning ties and outside ties. The last two types are related to cross-boundary technological exploration, but the mechanisms are different. We try to explain how the three types of formal agreements affect these two types of exploration.

Network theory developed two perspectives to analyse the effects of interorganizational networks on technological exploitation and exploration. The first view focusses on brokerage as the primary driver of technological exploration (Burt, 1992). The second view argues that technological exploration is driven by cohesiveness (Coleman, 1988). To propose how the three types of alliances affect the two types of technological exploration, we use the cohesiveness perspective for inside ties and the brokerage perspective for cross-spanning ties and outside ties. While cohesiveness promotes a more in-depth exploitative search, brokerage pursues a broader explorative search. However, explorative search has not been subcategorised before. The distinction between ELN and ELN has not been examined theoretically and empirically in prior studies. In our view, this opens up new ways in which the role of alliance partners in exploration may be considered. We argue that companies not only set up interorganizational ties to learn from their partners, but they also benefit from the “radar” function these partners play in reaching out to new, hitherto unknown technologies. This framework is empirically tested in the context of the application specific integrated circuits (ASIC) industry, a sub-sector of the semiconductor industry covering the entire population of ASIC producers in the period 1987–2000.

The contribution to the literature is threefold: first, we categorise two types of technological exploration, further refining the insights in the literature about explorative learning. Second, we consider that the role of partners is no longer restricted to the
co-development of new technologies, but they are also instrumental to a focal firm in reaching out to novel and useful sources of technologies. Third, different types of relations with partners have a different effect on the two types of exploration. Therefore, a firm has to carefully select the right partners when it intends to boost its explorative learning. In short, the contribution of our study lies not only in estimating the impact of these types of interorganizational relations on technological exploration, but also in the light it sheds on a novel distinction between two types of technological exploration that has not been investigated in previous research.

2. Theoretical background

2.1 The role of boundary-spanning technological exploration

Local search has been defined as the behaviour of organisations to search for solutions in the neighbourhood of their current expertise or knowledge (Stuart and Podolny, 1999; Phelps, 2010). Empirical evidence suggests that firms focus their technology search on closely related technological domains. By engaging in local search, firms can focus on similar technology and create incremental innovations (Laursen, 2012). And then firms can become even more expert in the technological domains they already master. Local search is beneficial when the competitive environment is stable and the technology dynamic is cumulative. However, core competencies can rapidly turn into core rigidities or fall into competency traps when new technologies emerge or when the competitive environment is changing rapidly (Leonard-Barton, 1992; Ahuja and Lampert, 2001; Kim et al., 2012). Under these circumstances, local search creates inertia, with the result that local search inhibits explorative search (Cyert and March, 1963; Hannan and Freeman, 1984). But, firms can overcome local search by deliberately repositioning themselves technologically. They can take action to explore emerging technologies and develop new technological capabilities in-house in order to secure long-term growth (Sirén et al., 2012).

Technological exploration cannot be achieved without searching activities beyond the boundaries of the firm (Rosenkopf and Nerkar, 2001; Rosenkopf and Padula, 2008). Exploratory search involves a conscious effort to look beyond the current knowledge base, in contrast with local search, where companies only use and extend their existing knowledge base (Katila and Ahuja, 2002; March, 1991; Laursen, 2012). Firms have to reach out for new technologies, because existing technological capabilities have a limited capacity to generate innovative products, leading inevitably to declining growth opportunities for the company. Firms that search systematically for externally developed knowledge have better access to new information and technology, and in this way they improve their capability to explore new technologies.

Previous research shows that firms must explore valuable knowledge which is developed by other organisations (Chesbrough, 2003; Laursen and Salter, 2006; Menon and Pfeffer, 2003; Rosenkopf and Nerkar, 2001). Many innovative firms have changed the way they search for new ideas, adopting open search strategies that involve the use of a wide range of external actors and sources to help them achieve and sustain innovation. The focus on tapping into external knowledge in studies of innovation reflects that the network of relationships between the firm and its external environment plays an increasingly important role in shaping performance.

Rosenkopf and Nerkar (2001) and Rosenkopf and McGrath (2011) also argue that exploration beyond organisational boundaries persistently leads to better innovation performance. Therefore, search beyond the organisational boundaries becomes a major determinant in explaining the performance differences between firms. In short, innovating firms have to get involved in technological exploration and they are encouraged to explore technology externally, because nowadays useful knowledge is widely distributed (Chesbrough, 2003; Bogers et al., 2018).
2.2 Two types of technological exploration

Firms have to learn from other organisations to move beyond local search and explore new technological opportunities. But who are these organisations from which firms can learn? A broad stream of literature has focussed on technology alliances as a major conduit to source external technology or to co-develop technology with other organisations with complementary skills and capabilities (Hagedoorn, 1993; Gulati, 1995a, b; Podolny, 1994; Gulati et al., 2009). Although the literature has emphasised the role of technology alliances in explorative learning, companies have several options to explore new technologies. Firms can learn, for example, from scientific publications (McMillan et al., 2000), patent releases, contact with consultants, technology providers and intermediaries (Howells, 2006), product introductions in the market, conferences, exhibitions, benchmarking with competitors (Wiersema and Bowen, 2008; Hunt and Morgan, 1997), crowds (Afuah and Tucci, 2012, 2013), mobility of personnel, etc. These are just a few examples to illustrate that the establishment of formal interorganizational ties is only one way – although an important one – to source new technologies. In our view, the literature has underemphasised the role of these alternative ways to source new technologies. Therefore, we make a distinction between two types of explorative learning.

We have labelled the first type “ELP”. In ELP, a firm explores new technologies through establishing interorganizational relations with different partners. These ties are a formalised way of sourcing and co-developing technological knowledge. Most of these ties are alliances: some are equity-based, others focus on contractual agreements between partners. Recently, new forms of formal innovation relationships have been developed, such as accelerators, incubators, use of knowledge intermediaries/brokers, etc. (Weiblen and Chesbrough, 2015). Second, a firm can learn from organisations with which it has not established alliance agreements. We have labelled this second type “ELN”. ELN refers to organisations on whose technologies a focal firm can build its new technological capabilities without having established any interorganizational ties in the past.

We thus differentiate between two types of technological exploration. ELP or explorative learning through alliance partners has been studied extensively in the literature. In contrast, ELN has not captured the attention of researchers so far. In this study, we show that ELN is important and should be analysed in conjunction with ELP, in order to understand how firms actually rely on other organisations’ knowledge to explore new technologies.

The distinction between ELN and ELP also allows us to disentangle two different roles of technology partners in explorative learning. First, we will look at the role of these partners in ELP. Interorganizational relations are usually examined as a conduit of knowledge. The knowledge base of the partners to whom a company has access to is the main reason why technology partnerships are established. A technology agreement between two firms provides a reliable channel through which each partner can learn about the competences and the trustworthiness of the other (Rosenkopf and Padula, 2008). Thus, technology agreements are interesting instruments to facilitate explorative learning. Explorative learning entails both exogenous uncertainty (technological and market uncertainties) and endogenous uncertainty (opportunistic behaviour of the alliance partner) (Van de Vrande et al., 2006). And alliance agreements can be shaped in a way that minimises these risks. Over time, alliance experience can generate trust between partners (Gulati, 1995b). Trust reduces transaction cost and uncertainties involved in information sharing and transfer (Dyer and Chu, 2003; Li et al., 2008). Overall, alliance experience or repeated ties between partners is a mixed blessing in technological exploration. On the one hand, formal technology relations between organisations can reduce cost and uncertainties; on the other hand, a focal firm may not learn novel technologies from partners with whom it has already set up several technology projects in the past.
Interorganizational relations have the second role in explorative learning. This role has not been established in the literature and is related to ELN. In ELN, technology partners may foster exploratory learning by connecting a focal firm with interesting sources of technology. They thus act as a reputation reference or inform the focal firm about new technological opportunities. Thus, alliance partners play the role of a radar to discover new technological opportunities, as well as a reputation reference to facilitate the contacts between the focal firm and the potentially interesting sources of new technologies (Podolny, 2001; Gulati, 1998). As a result, the function of networks provides a vehicle for gathering information about potential partners through effective referrals (Burt, 1992; Rosenkopf and Padula, 2008). Due to their contacts with different technology providers, partners can provide the focal firm with social cues about the reliability of potentially interesting technology sources in other organisations that are not part of the formal network, thereby reducing the search costs and risks of exposure to opportunistic behaviour (Rosenkopf and Padula, 2008).

So, partners can also act as referrals or reputation references. Podolny (2001) argues that “the presence or absence of a tie between two market actors is an information cue on which others rely, in order to make inferences about the underlying quality of one or both of the market actors” (p. 35). When a focal firm has partners with strong technologies and a solid reputation in the industry, these can act as a referral for the focal firm among other organisations with interesting technologies, or they can facilitate the contact between the latter and the focal firm.

2.3 Interorganizational relations and technological exploration

Why and how do interorganizational ties have an impact on innovative firms’ technological exploration? Many previous studies have shown the positive impact of these ties on technological innovation and exploration (e.g. Ahuja, 2000; Capaldo, 2007; Powell et al., 1996; Belderbos et al., 2010; Garriga et al., 2013). These scholars argue that innovative firms sustain their innovative performance by tapping into the overall architecture of their formal collaborations. Interorganizational relations are structured as small worlds, characterised by clusters of locally embedded firms connected by a handful of “shortcuts” (Watts and Strogatz, 1998; Watts, 1999; Gulati et al., 2013). Clusters are characterised by strong cohesion between partners, and linking with partners within a cluster will be useful in order to reach information in a familiar context. For technological exploration, however, companies have to establish formal ties with partners outside the clusters to tap into new and unfamiliar technological knowledge. Because cluster-spanning ties serve as bridges spanning the structural holes across clusters, their use suggests access to less familiar contexts compared to that of prospective partners residing within the same cluster (Burt, 2005). Therefore, ties crossing the cluster boundaries are essential in explaining explorative learning. Consequently, boundary-spanning explorative learning should be analysed in the context of the small world characteristics of interorganizational ties, in which ties that cross-cluster boundaries play a crucial role (see also Baum et al., 2003).

As a result, we study the impact of alliances on technological exploration by focussing on the role of cliques in these alliances. Clique membership provides clique members with benefits, but creating links across cliques is also advantageous for the performance of clique members (see, e.g. Rowley et al., 2004, 2005; Shipilov, 2005).
To explore explorative search, we examine how clique members can benefit from different types of alliances. We categorise interorganizational ties in relation to their role vis-à-vis the clique to which the clique members belong. We distinguish between three types of ties: “inside ties” with other firms which are members of the same clique, “clique-spanning ties” linking the firm to partners in other cliques and “outside ties” linking the firm with “peripheral” firms. Peripheral firms are companies that do not belong to a clique and they are usually found in the periphery of a network, with few connections to companies in the rest of the network. Thus, clique members can reach across the boundary of the clique in two ways: firstly, “clique-spanning ties” make a bridge to partners in another clique. Second, “outside ties” link a clique member to firms which are not a member of a clique themselves.

3. Hypotheses
We use two complementary views of social network theory cohesion and brokerage. Coleman (1988) examined the advantages of cohesion and redundant ties, while Burt (1992) emphasised the advantages of brokerage and bridging ties in the networks. While brokerage provides access to a wider variety of knowledge in stimulating new value creation, cohesiveness helps to integrate this knowledge to deliver value and generate innovation (Capaldo, 2007; Gargiulo and Benassi, 2000; Padula, 2008; Tiwana, 2007). On the one hand, the advantages of inside ties are based on the cohesiveness of the clique. On the other hand, the advantages of clique-spanning ties and outside ties are based on brokerage in the alliances. Cohesive structures promote a deeper search of technological areas with which a firm is already familiar, since with inside ties it connects to firms that have ties with the existing alliance partners of the focal firm. Inside ties are assumed to be instrumental for technological exploitation. In contrast, bridging structures such as cross-spanning ties and outside ties bring the focal firm into contact with partners who do not usually have any ties with the existing alliance partners of the focal company themselves. In this way, we expect that clique-spanning ties and outside ties might be instrumental for firms which pursue a broader search strategy to explore new technologies. Having made a distinction between ELP and ELN in the previous section, we argue that these three types of alliances should have the different impacts on both types of technological exploration.

3.1 Inside ties and explorative learning
Inside ties are ties between firms in the same clique. Alliance cliques are characterised by strong and repeated ties within cohesive networks (Burt, 1992). Highly cohesive ties with many connections linking one partner in the alliance to another are said to improve the innovativeness of the alliance members (Coleman, 1988; Ahuja, 2000; Lazzarini, 2007; Rowley et al., 2004; Guler and Nerkar, 2012; Padula, 2008). Over time, ties inside existing cohesive networks exert a positive impact on the innovative capability of the firms, since they create trust and reciprocity norms that facilitate knowledge sharing between the clique members. However, inside ties also generate isomorphism between firms within the clique, thereby decreasing network diversity and firms’ access to non-redundant knowledge (Burt, 1992). In other words, inside ties may be useful to develop and refine the current knowledge base of a firm, but they are rather a liability for explorative learning, since they limit firms’ access to new or non-redundant information. A newly established inside tie delivers redundant information, because the focal firm was already connected to alliance partners of the new partner. The dense network of ties provides many redundant paths to the same nodes and the same sources of information (Burt, 1992; Granovetter, 1973). Therefore, we expect that inside ties will not be instrumental in invigorating explorative learning among clique members. Hence, we argue that inside ties have a negligible effect on ELP.
Similarly, there are reasons to assume that inside ties have no beneficial effects on ELN. In ELN, the partner should enable the focal firm to come into contact with new external sources of technology. However, inside ties reinforce the tendency to stick with embedded partners, who have similar information about new technologies and about other technology sources as the other partners in the clique. Thus, inside ties tend to insulate firms and prevent them from exploring beyond their existing networks (Capaldo, 2007; Padula, 2008; Gulati, 1995b; Uzzi, 1997). In sum, inside ties reduce the scope of explorative search. Hence, we hypothesise:

\( H1a. \) Inside ties have a negligible effect on both ELP.

\( H1b. \) Inside ties have a negligible effect on both ELN.

3.2 Clique-spanning ties and explorative learning

Clique-spanning ties connect a firm belonging to a clique to a partner who is a member of another clique. A clique-spanning tie is thus a bridging tie between two cliques (Burt, 1992), spanning a structural hole that separates two cliques. In this way, a firm can access less familiar technologies compared to those developed by partners who reside within the same clique (Burt, 2005; Rosenkopf and Padula, 2008; Cowan and Jonard, 2009). Clique-spanning alliances can create the brokerage advantages. Only a few competitors have the same type of ties, which, in turn, generates a competitive advantage for the focal firm (Burt, 2005; Rosenkopf and Padula, 2008). Therefore, clique-spanning ties provide a reliable channel through which each partner can learn about new technologies. As the heterogeneity of alliance partners increases, the range and diversity of ideas, perspectives and information increase too. Clique-spanning ties expose firms to novel alternatives and broader, varied knowledge that enables them to explore more and generate breakthrough innovations (Katila and Ahuja, 2002; March, 1991; Cowan and Jonard, 2009).

Also, clique-spanning ties have another advantage. In a clique-spanning tie, the partner is acquainted with the technology of its own partners in the clique, which implies that the focal firm not only can learn about the technology of its partner but also about the knowledge generated by its partner’s partners in the same clique. Clique-spanning alliances create brokerage advantages (Burt, 2005; Rosenkopf and Padula, 2008). By definition, only a few competitors will have the same type of ties, and this, in turn, generates a competitive advantage for the focal firm. Therefore, clique-spanning ties are instrumental in improving ELP.

Do clique-spanning ties also facilitate ELN? In other words, do partners bring the focal firm into contact with new interesting technology sources? As mentioned earlier, a clique member is highly embedded in its own clique. Highly embedded firms risk generating a cognitive lock-in, which isolates it from the outside world (Capaldo, 2007; Padula, 2008). The strong embeddedness within its clique of partners will prevent it from having a broad overview of the technologies beyond the clique (Uzzi, 1997). The focus of the partner on its partners in the same clique prevents it from staying in continuous touch with other or novel sources of technology. Consequently, we argue that clique-spanning ties will not have a significant impact on ELN. Hence, we hypothesise:

\( H2a. \) Clique-spanning ties have a positive effect on ELP.

\( H2b. \) Clique-spanning ties have no effect on ELN.

3.3 Outside ties and explorative learning

Outside ties connect a focal firm to non-redundant contacts with unique information that inside ties and clique-spanning ties cannot provide. The partners in outside ties have a
peripheral position in the network of ties. Their knowledge is, by definition, non-redundant. Moreover, they have a peripheral position because they are new in the industry or they have developed knowledge about alternative or emerging technologies. Therefore, outside ties are beneficial for firms’ ELP, because they are conduits to new, unique and heterogeneous information (Granovetter, 1973; Hansen, 1999). However, although the information from outside ties may be very valuable, partners in outside ties are not embedded in a rich set of ties themselves. Hence, an outside tie will provide information about the technology of the partner, but since the partner is poorly connected to other organisations it will not incorporate knowledge from these partners (as in the case of cross-spanning ties). Therefore, we argue that outside ties will have only a moderate positive effect on ELP.

Alliances with peripheral firms also give a firm access to a wide spectrum of new information and help to deal with technological dynamics (Lin et al., 2009). In outside ties, the partner is peripheral in the network and can, therefore, bring the focal company into contact with new or unfamiliar sources of information and technology. In contrast with clique-spanning ties, the partner is not hindered through its clique-embeddedness from staying in touch with other remote sources of technology. Therefore, we expect that outside ties promote ELN by bringing the focal firm into contact with a wider range of technological knowledge. Outside ties provide a vehicle for gathering information about technology sources through effective referrals by the partner in knowledge fields with which the focal firm is not familiar. Hence, we hypothesise:

**H3a.** Outside ties have a moderately positive effect on ELP.

**H3b.** Outside ties have a positive effect on ELN.

### 3.4 Technological capital and explorative learning

To understand a firm’s explorative learning, we should also focus on the technological capital of the focal firm. In explorative learning, a firm should have enough absorptive capacity to access, assimilate and integrate external technological knowledge (Cohen and Levinthal, 1990; Lane and Lubatkin, 1998; Fabricio, 2009). For acquiring the external knowledge successfully, a firm needs to possess prior knowledge that facilitates the absorption of external knowledge. Absorptive capacity is crucial for explorative learning, since the knowledge by definition is relatively new for the firm. A firm’s technological capital can help to recognise relevant external knowledge sources and assimilate that knowledge more easily (Cohen and Levinthal, 1990). We expect that this is the case for both types of technological exploration. Hence, it is hypothesized that:

**H4a.** An increase in the technological capital (of a focal firm) will have a positive effect on ELP.

**H4b.** An increase in the technological capital (of a focal firm) will have a positive effect on ELN.

Technological capital plays a dual role in relation to explorative learning. On the one hand, a firm’s technological capital enhances its explorative learning, since the firm has a stronger absorptive capacity to source knowledge from outside sources and use this knowledge to generate new product and process innovations. On the other hand, technological capital also negatively moderates the impact of inter-firm ties on ELP. Interorganizational ties may facilitate access to complementary resources that are needed to explore new technologies. Companies can team up with other companies to monitor and stay in touch with the latest technological development as well as to support the transfer and integration of external knowledge. There is, however, a possible substitution effect between technological alliances as a major organisational mode to acquire external and
internal knowledge of these technologies. Companies with strong technological capital may not profit as much from their partners as firms with less technological capital. In the case of a firm with strongly technological capital, there is a higher probability that the ties with its partners will not lead to explorative learning, but rather exploitative learning. This is because a technologically savvy firm will have already developed some technological expertise in a particular technological field. In contrast, in the case of companies with smaller technological capital, there is a higher probability that they will increase their explorative learning, since interorganizational ties can help them to overcome their internal technological weaknesses.

In other words, larger technological capital can reflect the diversity of technologies in which the company is involved (especially for large multidivisional companies), while smaller technological capital can reflect the specialisation of a firm in a particular technology. Generally speaking, the specialised firm has a greater opportunity than the diversified company to explore new technologies through its alliances. As a result, the larger the technological capitals, the lower the impacts of alliances on ELP.

On the contrary, technological capital does not play a similar role in ELN. We do not expect that technological capital influences ELN. Here, partners play the role of referrals and “radars” to find new technology sources. A focal firm can benefit from this radar and referral function of its partners, and this effect will not be influenced by the strength of its technological capitals. Firms with either stronger or weaker technological capitals can benefit from being made aware of technology sources that they were not familiar with before.

We have no a priori reason to assume that these hypothesised effects of technological capabilities on ELP and ELN differ among the three types of alliances. Thus, we hypothesise:

\[ H5a. \] Technological capital negatively moderates the effect of the three types of ties on ELP.

\[ H5b. \] Technological capital does not affect the effect of the three types of ties on ELN.

4. Empirical setting
Our hypotheses are tested on the population of ASIC producers that were active in the period 1987–2000. ASICs are a special type of integrated circuit (IC) that accounted for about 12 per cent of worldwide IC sales in 1999. The ASIC market is a typical high-tech industry in which technology is the driving force, shaping competition among firms and R&D-to-sales ratios are exceptionally high. In contrast with the general purpose ICs such as DRAMs and microprocessors, ASICs are built to perform only one particular function, e.g. converting digital signals of a CD or MP3 file into music. ASIC technology has been around since the early 1970s, but the industry began to show strong development in the early 1980s, when various electronic devices such as the desktop computer and the microprocessor became a success. IC customers started to realise that a good custom-designed IC could give them a sustainable competitive advantage over their competitors, and they were willing to pay for IC design together with the fixed costs of IC production. The industry started out with ASIC products offered by large integrated IC companies or produced in-house by large manufacturers of electronic devices (e.g. IBM and Texas Instruments). The increasing complexity and specialisation of ASIC products initiated a process of start-ups and spin-offs between 1980 and 1990, with companies focussing exclusively on the ASIC industry.

The ASIC market is divided into three segments: semi-custom ASICs, such as gate arrays and linear arrays; custom ASICs, such as full custom and standard cells; and programmable logic devices (PLDs), such as field programmable gate arrays (FPGA) and electronically programmable analogue circuits (EPAC). Formal definitions are given in the ASIC definitions provided below (Source: ICE ASIC-Outlook industry reports, McLean, 1987–2000) and
diagrammed in Figure 1. Firms may be involved in just one of these segments, or in more segments at the same time. Segments are important in the sense that firms in each segment face different competitors, technologies and competitive or technological dynamics. Customers typically make a decision between the three ASIC segments based on the total cost per chip, which in turn is dependent on the production volume and the design complexity. PLDs are the cheapest solution for simple and low volume production, while full custom designs are the most efficient solution for production volumes that exceed several hundred thousand ASICs.

ASIC definitions:

1. Semi-custom IC: a monolithic circuit that has one or more customised mask layers, but does not have all mask layers customised, and is sold to only one customer:
   - Gate arrays: a monolithic IC usually composed of columns and rows of transistors. One or more layers of metal interconnect and are used to customise the chip.
   - Linear array: an array of transistors and resistors that performs the functions of several linear ICs and discrete devices.

2. Custom IC: a monolithic circuit that is customised on all mask layers and is sold to only one customer:
   - Standard cell IC: a monolithic circuit that is customised on all mask layers using a cell library that embodies pre-characterised circuit structures.
   - Full custom IC: a monolithic circuit that is at least partially “handcrafted”. Handcrafting refers to custom layout and connection work that is accomplished without the aid of standard cells.

3. PLD: a monolithic circuit with fuse, antifuse or memory cell-based logic that may be programmed (customised), and in some cases, reprogrammed by the user:
   - FPGA: a PLD that offers fully flexible interconnects, fully flexible logic arrays and requires functional placement and routing.
   - EPAC: a PLD that allows the user to programme and reprogram basic analogue devices.

The development and production of ASICs requires interplay between various economic agents. The most important participants are the ASIC design houses, IC manufacturing
facilities, electronic system manufacturers and CAD tool vendors. To this list, we might add a number of auxiliary and/or intermediate players, such as companies offering services in the microelectronics field, firms that translate customers’ needs into the specifications for the design of ASICs and university labs. Some large system manufacturers have their own ASIC design house and foundry, or they acquire one, but even in this case they still cooperate with specialist design houses on account of recurrent peaks in design work. Large electronic system manufacturers seek to gain a foothold in the ASIC market by vertical integration: they aim to achieve a competitive advantage for their electronic systems through proprietary ASIC designs. These large, integrated electronic system manufacturers usually have their own fablines and their ASICs are processed together with standard ICs. They also make corporate-wide deals and second-source agreements with foundries. Smaller electronic companies set up agreements with different foundries and vendors to design and process their ASICs. As ASIC designs become increasingly complex, companies establish numerous joint development and cross-licensing agreements. Given these characteristics of the industry, most strategic alliances in the ASIC industry are likely to be strategic tools for external technology sourcing or joint development. In a high-tech environment like the ASIC industry, firms establish strategic alliances with each other in order to keep up with the newest technologies (Duysters and Hagedoorn, 1996), making this an interesting industry for testing our research questions.

5. Data and variables

5.1 Data

We constructed a panel data set covering the population of ASIC producers over the period 1987–2000. Based on the vendor list included in the ICE ASIC-Outlook industry reports (McClean, 1987–2000), we established a detailed list of all ASIC producers. In order to measure the technological knowledge base of the ASIC producers, we draw on patent data from the US Patent and Trademark Office. In industries like the ASIC industry in which companies operate on a global scale, US patents are a good proxy for companies’ worldwide technological performance and technological assets. The data on strategic technology alliances were obtained from the ICE industry reports; the ASIC-Outlook reports (McClean, 1987–2000) and the MERIT–CATI database on strategic technology alliances (Hagedoorn, 1993). Financial data about ASIC producers have been gathered from different sources, including the annual ICE reports (McClean, 1987–2000) and COMPUSTAT.

5.2 Alliance networks and operationalisation of cliques

In order to differentiate between the three types of ties, first we have to identify cliques in the network. For each year in the period 1987–2000, we constructed the network in the ASIC industry based on the alliances that were established during the five years prior to the year under observation. Based on these networks, we constructed the “cliques”. In the construction of the cliques, we needed to formulate a measuring technique for grouping geodesically close firms that also guaranteed a density within the cliques that was higher than a lower bound threshold. Therefore, and in line with prior research on cliques (Rowley et al., 2005), we used the N-Clan procedure implemented in UCINET to detect relevant cliques (Borgatti et al., 2002). We defined the maximum path length between all partners in a clique as 2. The N-Clan procedure allows firms to be a member of more than one clique. In this particular industry, 38 per cent of the firms were members of more than one clique. In accordance with Rowley et al. (2005), firms assigned to multiple cliques were considered to be the members of each clique for the purpose of computing the clique-level independent variables. We corrected the firm-level observations for overrepresentation by weighting these observations by the number of cliques these firms participated in. We identified 75
active ASIC producer firms that were located within a clique during the period 1987–2000. This resulted in an unbalanced panel of 643 observations. In total, we found an average of 33 cliques per year during our observation period (1987–2000)[1].

5.3 Variables

*Dependent variables.* We make a distinction between two types of dependent variables: ELP and ELN. We refer to Figure 2 to explain the distinction between both variables. This figure shows how different types of learning can be distinguished by tracking the backward citations of new patents of an innovating firm in a particular year (following Rosenkopf and Nerkar, 2001). When companies build on prior technological knowledge, new patents must cite existing patents on which they build. As a result, patent citations are a unique and reliable instrument to define different types of exploration[2].

In some cases, firms invent new technologies that do not build on any prior art. These “pioneering technologies” have no technological antecedents and they represent the technologies that do not build on any existing technologies (Ahuja and Lampert, 2001). We can further distinguish two different cases when a new patent cites prior patents. On the one hand, a new patent can cite some of the firm’s own patents. Self-citations imply that the new patent is built on the firm’s prior expertise and experience. This type of patent represents a firm’s exploitative learning (Benner and Tushman, 2002; Rosenkopf and Nerkar, 2001; Schildt et al., 2005). Through exploitation, firms extend and refine their existing technical capabilities that are most likely to be the competencies that are crucial for the competitiveness of their current businesses. In other cases, firms can also successfully file new patents that do not cite any of their own prior patents. When a firm’s new patents have no backward self-citations, the firm explores new technological areas and broadens its own technological capabilities by building on the knowledge of other organisations. Patents with no self-citations but which cite patents from other firms are considered to be more explorative than those that also cite own prior technology. These patents are important for understanding different types of interorganizational ties.

![Figure 2. How to distinguish between the two types of explorative learning](image-url)
the purpose of avoiding problems related to local search (March and Simon, 1958; Nelson and Winter, 1982; Helfat, 1994; Eriksson et al., 2016).

So far, we have been using existing definitions of technological exploitation and exploration (Rosenkopf and Nerkar, 2001). We further segment exploratory patents into two subcategories. On the one hand, innovating firms may avoid local search by learning from their partners. A patent is categorised as ELP when there are no self-citations and when some backward citations refer to organisations with which the innovating firm has established one or more alliances during the last five years (Benner and Tushman, 2002; Schildt et al., 2005). In contrast, a patent is categorised as ELN when there are no self-citations and when there are only backward citations referring to organisations with which the innovating firm established no alliances during the last five years. This moving window approach is considered to be an appropriate timeframe during which the existing portfolio of strategic alliances is likely to have an influence on the current technological performance of a firm (Kogut, 1988, 1989; Gulati, 1995b).

Both dependent variables are count variables, indicating the number of new patent applications that fulfil the requirements of ELP or ELN. ELP is calculated as the sum of patents successfully applied for by the focal firm in each year, which have at least one citation to its alliance partners' prior patents, but no citations to its own prior patents. ELN is calculated as the number of patents successfully applied for per year by the focal firm which neither cites its own prior patents nor its alliance partners' patents.

Independent variables. The independent variables are based on the alliance networks in the ASIC industry and the cliques defined above. The cliques are derived from the alliances established during the five years prior to the year of observation. Companies can change their position in a clique by establishing new ties, and dissolving, strengthening or weakening existing ones (Koka et al., 2006). The new ties that are established by clique members can be classified as “inside clique ties”, “outside clique ties” and “clique-spanning ties”. If a newly established alliance in year \( t \) bridges two distinct cliques in the existing alliance network, then it can be considered a clique-spanning tie. Inside ties link two firms within the same clique and outside ties link a clique member to a firm which is not a member of a clique (we also label these partners as peripheral network members, since they are not part of the main pack in the alliance network). After categorising these ties, the number of clique-spanning ties can be measured by counting the number of clique-spanning ties that a firm initiated five years prior to the observation year.

Control variables. We include three organisational variables and four dummy variables to control for unobserved effects. In order to control for unobserved heterogeneity at the firm level, we included the variable “technological capital”, representing the total size of a firm's technological knowledge base. This variable was created by adding up all ASIC-related patents that a firm had successfully applied for during the five years prior to the year of observation. A moving window of five years is considered to be an appropriate timeframe for assessing technological impact in high-tech industries (Podolny and Stuart, 1995; Henderson and Cockburn, 1996). The second firm-level variable relates to the “size of the firm”. Large firms have a broader and more diversified established network of alliances (Hagedoorn and Duysters, 2002) and position themselves as dominant firms, not only within the clique, but also in the overall alliance network. Due to their size advantage, large firms are more likely to profit from economies of scale and scope, and so they have a higher potential to increase their technological performance over time. We calculated this variable as the natural logarithm of a firm's annual sales. The third firm-level control variable is “R&D intensity”. Firms that invest more in R&D have more options to explore and experiment with new kinds of technologies. We calculated this variable based on the R&D-to-sales ratio of each firm.
Furthermore, we include four types of dummy variables to control for different types of contingencies. A first dummy variable was included to control for a potential bias, since some large companies only produce ASICs to cover their internal needs (captive sales). These captive producers represent a small minority of ASIC-producing companies, but they are nonetheless important in terms of technological capabilities; consequently, they play an important role in the technological development of the ASIC industry (e.g. IBM and Texas Instruments). A second industry dummy variable is included to indicate the industry to which an ASIC producer belongs. Firms can be involved in the production of just one segment or in more segments of the ASIC industry at the same time. Segments are important in the sense that firms in each segment face different competitors, technological challenges and competitive or technological dynamics. The third dummy variable indicates the economic region in which the company is headquartered (Asia, North America or Europe), the default being North America (Ohmae, 1985). Finally, year dummy variables are included to capture changes over time in the propensity of firms to patent their inventions (Table I).

6. Results
In Tables II and III, we present the descriptive statistics and the correlation matrix for the different variables. Our dependent variables, ELP and ELN, are count variables. Because our data show evidence of over dispersion, a negative binomial regression model is the most

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Variable description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explorative learning from partners (ELP)</td>
<td>The number of patents successfully applied for in year t by the focal firm, which have at least one backward citation to its alliance partners’ prior patents, but no citations to its own prior patents</td>
</tr>
<tr>
<td>Explorative learning from non-partners (ELN)</td>
<td>The number of patents successfully applied for in year t by the focal firm, which neither cites its own prior patents nor its alliance partners’ prior patents</td>
</tr>
<tr>
<td>Prior clique-spanning ties (CST)</td>
<td>Alliances linking two firms belonging to two different alliance cliques using the prior alliance network in the ASIC industry (t−1 to t−5)</td>
</tr>
<tr>
<td>Prior outside ties (OT)</td>
<td>Alliances linking a firm belonging to an alliance clique with one that is not a clique member using the prior alliance network in the ASIC industry (t−1 to t−5)</td>
</tr>
<tr>
<td>Prior inside ties (IT)</td>
<td>Alliances linking two firms belonging to the same alliance clique – using the prior alliance network in the ASIC industry (t−1 to t−5)</td>
</tr>
<tr>
<td>Technological capital</td>
<td>Count variable indicating the number of successful patent applications in (t−1 to t−5)</td>
</tr>
<tr>
<td>Technological distance</td>
<td>Distance between technology portfolio of the focal firms and that of its partners (t−1 to t−5)</td>
</tr>
<tr>
<td>Firm Size</td>
<td>Total overall sales of the focal firm/1,000 (t−1)</td>
</tr>
<tr>
<td>Firm R&amp;D to sales ratio</td>
<td>Firm total R&amp;D expenditures/Firm’s overall sales (t−1)</td>
</tr>
<tr>
<td>Firm is captive producer</td>
<td>Dummy variable denoting that the firm is not selling products on the ASIC market</td>
</tr>
<tr>
<td>Firm is SC producer</td>
<td>Dummy variable denoting that the firm is only producing standard cells</td>
</tr>
<tr>
<td>Firm is PLD producer</td>
<td>Dummy variable denoting that the firm is only producing PLDs</td>
</tr>
<tr>
<td>Firm is GA and SC producer</td>
<td>Dummy variable denoting that the firm is only producing gate arrays and standard cells</td>
</tr>
<tr>
<td>Firm is GA and PLD producer</td>
<td>Dummy variable denoting that the firm is only producing gate arrays and PLDs</td>
</tr>
<tr>
<td>Firm is SC and PLD producer</td>
<td>Dummy variable denoting that the firm is only producing standard cells and PLDs</td>
</tr>
<tr>
<td>Firm is GA and SC and PLD producer</td>
<td>Dummy variable denoting that the firm is producing gate arrays and standard cells and PLDs</td>
</tr>
<tr>
<td>Firm is European</td>
<td>Dummy variable denoting that the firm is headquartered in Europe</td>
</tr>
<tr>
<td>Firm is Asian</td>
<td>Dummy variable denoting that the firm is headquartered in Asia</td>
</tr>
<tr>
<td>Dummy 1987–1999</td>
<td>Dummy variable denoting the year of observation</td>
</tr>
</tbody>
</table>

Table I. Definitions of dependent and independent variables.
appropriate method (Cameron and Triverdi, 1998). To determine the choice between a random-effect and fixed-effect model, we conducted a Hausman (1978) test. The Hausman test indicated that random effects estimators are consistent and efficient for this analysis.

Table III presents the correlations between the different variables. Except for correlations between main variables and their interaction terms (such as the combinations of the alliance types and technical capital), all correlations among the independent variables are low. However, cross-spanning ties and outside ties are positively related, and the R&D intensity is negatively correlated to the firm size measure (because many ASIC start-ups have an unusually high R&D intensity). The results of the regression diagnostics (e.g. VIF statistics) suggested, however, that multicollinearity was not a problem.

Table III shows the results of the random effects negative binomial regression analyses for 643 observations (of 80 clique members) in the sample. The left side of Table III represents the regressions for ELP, the right side the regressions for ELN.

Models 1 (ELP) and 7 (ELN) are the base model, only including control variables. Model 2 includes the three types of ties [4]. Inside ties do not have a significant impact on either type of explorative learning, providing empirical support for \( H1a \): firms partnering with other companies in the same clique will not benefit from these ties when they intend to explore...
<table>
<thead>
<tr>
<th></th>
<th>ELP</th>
<th>ELN</th>
<th>PCT</th>
<th>OT</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELP</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELN</td>
<td>0.3793</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCT</td>
<td>0.3369</td>
<td>0.1014</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OT</td>
<td>0.0565</td>
<td>0.0129</td>
<td>-0.0022</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>0.2169</td>
<td>-0.0234</td>
<td>0.4523</td>
<td>0.0885</td>
<td>1.0000</td>
</tr>
<tr>
<td>Tcap</td>
<td>0.4238</td>
<td>0.6412</td>
<td>0.2518</td>
<td>0.0317</td>
<td>0.0628</td>
</tr>
<tr>
<td>ln(sales)</td>
<td>0.1635</td>
<td>0.0623</td>
<td>0.2822</td>
<td>0.0777</td>
<td>0.2481</td>
</tr>
<tr>
<td>rd/sales</td>
<td>-0.0402</td>
<td>-0.0354</td>
<td>-0.0738</td>
<td>0.0004</td>
<td>-0.0674</td>
</tr>
<tr>
<td>captive</td>
<td>-0.0831</td>
<td>-0.0833</td>
<td>-0.0735</td>
<td>-0.0863</td>
<td>-0.0789</td>
</tr>
<tr>
<td>dga</td>
<td>-0.1279</td>
<td>-0.0963</td>
<td>-0.1990</td>
<td>-0.0653</td>
<td>-0.1725</td>
</tr>
<tr>
<td>dsc</td>
<td>-0.1293</td>
<td>-0.1410</td>
<td>-0.1440</td>
<td>-0.0351</td>
<td>-0.1522</td>
</tr>
<tr>
<td>dgasc</td>
<td>-0.0752</td>
<td>-0.0044</td>
<td>0.0729</td>
<td>-0.0159</td>
<td>0.0083</td>
</tr>
<tr>
<td>dscpld</td>
<td>0.0373</td>
<td>-0.0184</td>
<td>0.0989</td>
<td>0.0015</td>
<td>0.1371</td>
</tr>
<tr>
<td>dgapld</td>
<td>0.1111</td>
<td>0.0184</td>
<td>-0.0004</td>
<td>0.1764</td>
<td>0.0499</td>
</tr>
<tr>
<td>dgascpld</td>
<td>0.1213</td>
<td>-0.0245</td>
<td>0.2559</td>
<td>0.1381</td>
<td>0.3469</td>
</tr>
<tr>
<td>Eur firm</td>
<td>-0.0780</td>
<td>-0.1444</td>
<td>0.0554</td>
<td>0.1766</td>
<td>0.3944</td>
</tr>
<tr>
<td>Asia firm</td>
<td>-0.0268</td>
<td>0.1025</td>
<td>-0.0544</td>
<td>-0.0716</td>
<td>-0.1328</td>
</tr>
<tr>
<td>year 87</td>
<td>-0.0705</td>
<td>-0.0593</td>
<td>-0.1231</td>
<td>0.0891</td>
<td>-0.0428</td>
</tr>
<tr>
<td>year 88</td>
<td>-0.0734</td>
<td>-0.0506</td>
<td>-0.0746</td>
<td>0.1287</td>
<td>0.0094</td>
</tr>
<tr>
<td>year 89</td>
<td>0.0205</td>
<td>-0.0282</td>
<td>-0.0528</td>
<td>0.1237</td>
<td>0.0338</td>
</tr>
<tr>
<td>year 90</td>
<td>-0.0212</td>
<td>-0.0187</td>
<td>-0.0325</td>
<td>0.0999</td>
<td>0.0286</td>
</tr>
<tr>
<td>year 91</td>
<td>-0.0138</td>
<td>-0.0374</td>
<td>-0.0325</td>
<td>0.0541</td>
<td>0.0299</td>
</tr>
<tr>
<td>year 92</td>
<td>-0.0561</td>
<td>-0.0677</td>
<td>0.0208</td>
<td>-0.0096</td>
<td>-0.0105</td>
</tr>
<tr>
<td>year 93</td>
<td>0.0010</td>
<td>-0.0151</td>
<td>0.0312</td>
<td>-0.0543</td>
<td>0.0301</td>
</tr>
<tr>
<td>year 94</td>
<td>-0.0261</td>
<td>-0.0185</td>
<td>0.0361</td>
<td>-0.0749</td>
<td>0.0025</td>
</tr>
<tr>
<td>year 95</td>
<td>0.0589</td>
<td>0.0064</td>
<td>0.0444</td>
<td>-0.0666</td>
<td>0.0218</td>
</tr>
<tr>
<td>year 96</td>
<td>0.0513</td>
<td>-0.0084</td>
<td>0.0902</td>
<td>-0.0680</td>
<td>0.0332</td>
</tr>
<tr>
<td>year 97</td>
<td>0.0473</td>
<td>0.0479</td>
<td>0.0237</td>
<td>-0.0924</td>
<td>0.0223</td>
</tr>
<tr>
<td>year 98</td>
<td>0.0353</td>
<td>0.0685</td>
<td>-0.0073</td>
<td>-0.0385</td>
<td>-0.0546</td>
</tr>
<tr>
<td>year 99</td>
<td>0.0070</td>
<td>0.0896</td>
<td>0.0425</td>
<td>0.0346</td>
<td>-0.0644</td>
</tr>
<tr>
<td>cap</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(sales)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rd/sales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>captive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dga</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dsc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dgasc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dscpld</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dgapld</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dgascpld</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European firm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian firm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year 99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table III. Cross-correlation table (continued)
new technologies. In other words, clique members are not the right partners to explore new technologies. As Model 8 shows, nor are clique members of interest as referrals to learn about other relevant external technology sources, since the information these partners provide will be redundant information in most cases; the focal firm already has a lot of information about these sources through the existing partnerships in the clique. This finding supports H1b.

We expect a positive effect of clique-spanning ties on learning from partners. Model 2 in Table IV shows that clique-spanning ties have a significant positive impact on exploration from partners, which supports H2a. Similarly, we expect outside ties to have an effect on ELP as well. However, this is not the case: outside ties have a positive effect, but the coefficient is not significant. In other words, H3a is not supported by the data.
<table>
<thead>
<tr>
<th>Exploration from partners (ELP)</th>
<th>Exploration from non-partners (ELN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clique-spanning ties (CST)</td>
<td></td>
</tr>
<tr>
<td>CST×techn. capital</td>
<td></td>
</tr>
<tr>
<td>Outside ties (OT)</td>
<td></td>
</tr>
<tr>
<td>OT×techn. capital</td>
<td></td>
</tr>
<tr>
<td>Inside ties (IT)</td>
<td></td>
</tr>
<tr>
<td>IT×techn. capital</td>
<td></td>
</tr>
<tr>
<td>Technological capital/100</td>
<td></td>
</tr>
<tr>
<td>Firm size</td>
<td></td>
</tr>
<tr>
<td>R&amp;D intensity</td>
<td></td>
</tr>
<tr>
<td>Captive producer</td>
<td></td>
</tr>
<tr>
<td>GA producer</td>
<td></td>
</tr>
<tr>
<td>SC producer</td>
<td></td>
</tr>
<tr>
<td>GA and SC producer</td>
<td></td>
</tr>
<tr>
<td>SC and PLD producer</td>
<td></td>
</tr>
<tr>
<td>GA and PLD producer</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
Table IV.

<table>
<thead>
<tr>
<th></th>
<th>Exploration from partners (ELP)</th>
<th>Exploration from non-partners (ELN)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3) (4) (5) (6) (7)</td>
<td>(8) (9) (10) (11) (12)</td>
</tr>
<tr>
<td>GA, SC and PLD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>producer</td>
<td>−1.633*** −1.845*** −1.964***</td>
<td>−1.779*** −1.670*** −2.083*** −1.930*** −1.833*** −1.684*** −1.716*** −1.678*** −1.843***</td>
</tr>
<tr>
<td></td>
<td>(0.429) (0.440) (0.433) (0.453)</td>
<td>(0.446) (0.457) (0.430) (0.445) (0.430) (0.433) (0.435) (0.448)</td>
</tr>
<tr>
<td>Firm is European</td>
<td>−0.621 −0.533 −0.547 −0.716</td>
<td>−0.794* −0.795* −0.824** −0.850** −0.726** −0.834** −0.847** −1.0010**</td>
</tr>
<tr>
<td></td>
<td>(0.441) (0.427) (0.419) (0.453)</td>
<td>(0.432) (0.444) (0.330) (0.386) (0.368) (0.316) (0.386) (0.399)</td>
</tr>
<tr>
<td>Firm is Asian</td>
<td>−0.617 −0.370 −0.430 −0.670</td>
<td>−0.578 −0.382 0.375 0.617* 0.572* 0.499 0.552* 0.627**</td>
</tr>
<tr>
<td></td>
<td>(0.420) (0.400) (0.381) (0.433)</td>
<td>(0.425) (0.399) (0.330) (0.318) (0.316) (0.316) (0.314) (0.317)</td>
</tr>
<tr>
<td>Constant</td>
<td>−0.200 −0.206 0.683 0.036</td>
<td>0.158 0.577 −0.512 −0.539 −0.560 −0.548 −0.610 −0.451</td>
</tr>
<tr>
<td></td>
<td>(0.912) (0.886) (0.907) (0.942)</td>
<td>(0.948) (0.730) (0.741) (0.847) (0.748) (0.734) (0.745) (0.745)</td>
</tr>
<tr>
<td>Observations (firms)</td>
<td>643(80) 643(80) 643(80) 643(80)</td>
<td>643(80) 643(80) 643(80) 643(80) 643(80) 643(80) 643(80) 643(80) 643(80) 643(80)</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>−677.2 −671.5 −666.9 −675.2</td>
<td>−674.3 −662.1 −636.0 −625.8 −627.5 −626.7 −628.3 −624.3</td>
</tr>
</tbody>
</table>

**Notes:** This table shows the results of random effects negative binomial model. Standard deviations are shown in parentheses. Year dummy variables are included in the regressions but are not reported. *p < 0.10; **p < 0.05; ***p < 0.01
The opposite effect can be seen for ELN: here, outside ties have a positive effect on this type of explorative learning, while cross-spanning ties are no significant impact. Thus, we have strong empirical support for \(H2b\) and \(H3b\), while outside ties have no effect on ELP, they have a significant positive effect on ELN (see Model 8).

Models 2 and 7 in Table IV show that there is no immediate effect of technological capital on ELP. In contrast, it has a strong positive impact on ELN. Hence, we have support for \(H4b\), but not for \(H4a\). Models 3–5 in Table IV show that technological capital has no immediate effect on ELPs. In contrast, technological capital has a positive impact on ELN – see Models 9–11 in Table IV. We comment on this finding in the discussion section.

In \(H5a\) and \(H5b\), we suggested that there is a moderating effect of technological capital on the effect of the different types of ties on ELPs. Model 3–5 in Table IV show that technological capital negatively moderates the effect of the three types of alliances on ELP.

The combined effect of the ties and technological capital is represented in Figure 3. Firms with low technological capital profit from the three types of ties, although cross-spanning ties have a much bigger effect than the other two. Thus, firms with poorly developed technological capital can benefit not only from clique-spanning alliances, but also to a lesser extent from inside ties and outside ties. When their technological capital increases, then the effect becomes negative for outside ties and inside ties. This result shows that alliances should not be considered independently from technological capital in estimating their impact on ELP. In contrast, technological capital does not moderate the effect of the three types of ties on ELN. This sharp contrast between ELP and ELN – corroborating \(H5a\) and \(H5b\) – will be discussed further in the following section.

7. Discussion

This study investigates the effect of three types of interorganizational ties on two types of explorative learning. Firms can learn from their partners, as has been emphasised in prior studies. Firms with more partners will have more opportunities for explorative technologies, building on their knowledge. However, interorganizational ties are formal agreements to learn from external technology sources. In this study, we clarify that partners have two roles in technological exploration: a firm can learn from its partners because they have strong technological capabilities (ELP), or partners may function as a radar, bringing the focal firm into contact with new external sources of new or unexplored technologies (ELN). Furthermore, we analysed the effect of different types of alliances and internal technological capital on ELP and ELN.

7.1 Contributions to the theory

Four contributions emerge from the empirical analysis in this study.

Firstly, we contribute to the literature about technological exploitation and exploration. Only a few studies have shown the reasons why firms try to search beyond their boundaries (Rosenkopf and Nerkar, 2001; Rosenkopf and Padula, 2008). We propose that firms can explore new technologies beyond local search, but they can also do this in two different ways. This is one of the first studies to refine the concept of exploration by making a distinction between explorative learning, which focusses on the knowledge bases of the alliance partners, and learning through which a firm uses its technology partners to learn from organisations with which it has no formal technology agreements. The last type of exploration has not been studied empirically in the literature.

Second, interorganizational ties are important means for exploring new knowledge and technologies, since they allow companies to bridge technological domains effectively (Rosenkopf and Almeida, 2003). Our study started from the role of cliques: in order to source new technologies, a clique member has to go beyond the boundaries of its clique. We therefore distinguished three types of ties – inside ties, cross-spanning ties and outside ties.
ties– and we examined whether or not they are useful for boosting the two types of explorative learning in the ASIC industry. Our results show that inside ties have a negligible effect on both types of learning. A firm which is strongly embedded in a clique of partners will not experience any help from them in exploring new technological areas. Inside ties do not boost ELP, because the knowledge of an inside tie is redundant, as the new partner was already indirectly connected to the focal firm though other partners in the clique. Inside ties are not productive for learning from non-partners either (ELN), because the new partner will connect the focal firm with technology sources with which it is already familiar from other partnerships in the clique. Cross-spanning ties are expected to have a positive effect on ELP. This is fully supported by the results: when a firm can bridge two cliques by setting up a

Figure 3. Combined effect of the three types of alliances and technological capital on ELP

Notes: Horizontal axis represents technological capital. The four lines represent respectively the number of ties (0, mean, mean +1 SD; mean + 2 SD)
new alliance, it has a great source of technology to tap into. As expected, clique-spanning ties are not useful in advancing ELN: in clique-spanning ties, partners are heavily embedded in their own clique, preventing them to provide the focal firm with technology sources beyond their own alliance partners. Outside ties, in contrast, have no effect on ELP, but they have a positive effect on ELN: this implies that the focal firm is not directly exploring new technologies based on its partner’s knowledge base. Rather, it sets up an alliance with a partner to learn from new technology fields through the partnership: the partner is instrumental in informing the company about novel technologies and in facilitating access to them. In sum, this study shows that the choice of the right type of ties with partners in the network plays a crucial role in obtaining success with the two types of explorative learning.

Third, our study has research implications on the role of cohesiveness and brokerage in alliance networks (Burt, 1992; Coleman, 1988). A number of studies demonstrate that cohesiveness and brokerage play a complementary role in alliance networks in supporting firms’ innovation performance (Capaldo, 2007; Padula, 2008; Schilling and Phelps, 2007; Tiwana, 2007; Ahuja et al., 2012). By distinguishing three types of alliance ties, our study contributes to this line of literature by providing empirical evidence of how the cohesiveness and brokerage of the alliance networks are important in determining the success of the two types of technological exploration. Our results demonstrate that network closure has no effects on either type of technological exploration, while network brokerage shows different effects of cross-spanning ties and outside ties on the two types of technological exploration. Our results support two points: firstly, network closure (or inside ties) may play a positive role in exploitation, but it has no impact on technological exploration (Capaldo, 2007; Gargiulo et al., 2009); second, network brokerage has been identified as an influential determinant for enhancing innovative performance (Ahuja, 2000; Sytch et al., 2012; Zaheer and Bell, 2005). Our results specify that crossing the clique boundary with interorganizational ties leads to different results depending on the network position of the partnering firm. Linking up with firms in the periphery has a different effect compared to cross-spanning ties, where partners are members of a clique themselves.

Fourth, in addition to the main effects of different types of interorganizational ties, our study also shows how technological capital moderates the relationship between the three types of ties and the two types of technological exploration. The best way to study the effect of technological capital on ELP is to show the combined effect of a particular type of tie and technological capital: the results are summarised in the three graphs in Figure 3. The horizontal axis in these graphs represents the technological capital of ASIC producers (up to the mean + 2×SD). The different lines represent different values for the alliances (zero, mean, mean+SD and mean + 2×SD). The first graph shows that firms with few clique-spanning ties boost ELP when they have more technological capital. A high number of clique-spanning ties improve ELP drastically, but higher levels of technological capital reduce the effect on ELP. The effect of outside ties and inside ties on ELP is much smaller than in the case of clique-spanning ties. Moreover, relying on these two types of alliances may have a negative effect on ELP at higher levels of technological capital. This is certainly the case for inside ties, where negative values for ELP are recorded at moderate values of technological capital. In sum, relying on alliance partners is positive for companies with poorly developed technological capabilities (but the effect is much smaller for outside and inside ties). Companies with strong technical capabilities should avoid relying on inside and outside ties, while clique-spanning ties still have a positive effect. The effect of technological capital is straightforward in the case of ELN: companies with strong technological capabilities have a greater absorptive capacity to learn from new technology sources beyond their alliance partner network. There is – in contrast with ELP – no interaction between technological capital and the firms’ alliances.
7.2 Managerial implications
The current study has also some interesting implications for the management of interorganizational ties to maximise explorative learning. First, the concept of using technology partners as a prism (Podolny, 2001) is important and has not received the attention it deserves: technology partners are important for a focal firm because of the technology they own (ELP), but they are also valuable to detect and understand new technologies beyond their own technology (ELN). Technology partners facilitate the process for a company to learn about new technological developments. Second, given this dual role of partners in explorative learning, it is important to understand the position of the partners in the network. The three types of ties we discussed have different implications for both types of learning. Clique-spanning ties are useful for ELP, while outside ties – linking a firm to partners that are peripheral in the network – are instrumental for ELN. Third, managers have to keep in mind that the technological capital in their company has not any effect on the impact of the three types of ties on ELN. This is different for ELP: firms with strong technological capabilities experience less benefits from collaborating with technology partners and in the case of inside and outside ties, this may even have a negative effect on explorative learning.

7.3 Limitations and future directions
Our research is explorative and limited in several ways. Firstly, the operationalisation of the two types of explorative learning could be criticised. We have been using these concepts in an exclusive way. A patent that cites prior patents of partners is categorised as learning from partners, irrespective of the number of citations to non-partners. The analysis can be improved by developing more sophisticated, continuous variables that range between 100 per cent partner citations and 100 per cent non-partner citations. Second, the current study can be easily extended to explore both organisational and technological boundary-spanning activities (Rosenkopf and Nerkar, 2001). We have only focussed on exploration as an organisational boundary-spanning activity. Including technological boundary spanning would add complexity and would certainly enrich the analysis. Third, we used three types of interorganizational ties: the technology network literature has developed other categorisations of alliances that may be relevant when studying explorative learning. Finally, extending the types of relationships between partners (e.g. licensing, arm’s length R&D-contracting, patent search, informal/personal contacts, crowdsourcing, knowledge brokers, etc.) may also help to obtain a more accurate picture of how external sources of knowledge enhance explorative learning by firms.

Our investigation of the impact of different types of alliances on explorative learning provides new insights on inter-organisational learning. Technological exploration has received considerable attention in the literature, but many questions about how alliances can play a role in boosting explorative learning remain unanswered. We encourage other scholars to further engage in this interesting field of research.

Notes
1. The standard deviation is 5.16. The number of cliques varied over the years from 26 to 42.
2. We use patents and patent citations as a proxy for knowledge and knowledge flows, in line with the work of others (e.g. Ahuja and Katilla 2001; Almeida, 1996; Hoek and Agarwal, 2007; Jaffe et al., 1993; Peri, 2005). Knowledge may flow between individuals and firms through a number of mechanisms including conferences, publications, professional social networks and reverse engineering, in addition to patent reviews. Despite the various mechanisms for knowledge flow, knowledge and knowledge flows often leave their footprint in the form of patents (Jaffe, 1986). As a result, patents are an effective proxy for knowledge regardless of the mechanism. The fact that examiners add citations to patent applications is not really a concern here. Independent examiners
are involved in assessing the prior knowledge on which an innovation builds. The criticism may be made that these citations included by examiners do not reveal how a firm is building its technology on that of other companies. However, we feel comfortable that this is not a major issue for two reasons: firstly, as we mentioned earlier, learning is not necessarily based on explicit patent reviews. Secondly, should these citations be considered as random noise, then we can argue that if the empirical results in this paper reveal interesting relations, then they would be a priori hold in case we could eliminated the patent citations that the patent examiners added. Moreover, these extra citations lead to a better measure of the knowledge that an innovation actually builds upon.

3. “Patents successfully applied for” implies that these patents have been granted. However, we do not allocate a patent to the year it was granted, but prefer to allocate it to the year it was applied for, because at that time the invention had already been made.

4. We also ran the regressions introducing just one type of alliance into the regression. The results are very similar to those reported in Models 2 and 7.

References


Different types of interorganizational ties


**Further reading**

About the authors
Yu-Shan Su is Distinguished Professor of Technology Management at National Taiwan Normal University, Taipei, Taiwan. Her research interests are innovation management in the high-tech industry. She is Associate Editor of Technological Forecasting and Social Change. She is Editorial Review Board Member of Academy of Management Perspectives and Management and Organisation Review. She has published her works in Asia Pacific Journal of Management, Journal of World Business, Technological Forecasting and Social Change, International Journal of Technology Management, Tsinghua Management Review (Chinese), Sun-Yet Sen Management Review (Chinese), and others. She also co-authored and published a book of Innovation Management: Winning the Competitive Advantage in the Globe in Chinese in 2015. Yu-Shan Su is the corresponding author and can be contacted at: yssu@ntnu.edu.tw

Wim Vanhaverbeke is Professor of Strategy and Innovation Management at the Neoma Business School, France and is Visiting Professor at ESADE Business School and till recently also at the National University of Singapore. He published in different international journals such as Organisation Science, Research Policy, Journal of Product Innovation Management, California Management Review, Journal of Management, Journal of Management Studies, Small Business Economics, Journal of Business Venturing, Technovation, etc. He was co-editor with Henry Chesbrough and Joel West of Open Innovation: Researching a New Paradigm (OUP, 2006) and New Frontiers in Open Innovation (OUP, 2014). He published a management book “Managing Open Innovation in SMEs (CUP, 2017) and an academic volume Researching Open Innovation in SMEs (World Scientific Press, 2018). He was recognised by the International Association of Management of Technology (IAMOT) as one of the top 50 authors of technology and innovation management during the period 2008–2012. He was appointed as Member of the Advisory Committee of the Research Centre for Technological Innovation of the Tsinghua University from April 2018 to March 2021.