Journal of Business Research 98 (2019) 261–276

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Big data analytics and firm performance: Findings from a mixed-method approach

Patrick Mikalef[a,*], Maria Boura[b], George Lekakos[b], John Krogstie[a]

[a] Department of Computer Science, Norwegian University of Science and Technology, Sem Sælandsvei 9, 7491 Trondheim, Norway
[b] Department of Management Science and Technology, Athens University of Economics and Business, 76 Patission Str. 10434 Athens, Greece

ARTICLE INFO

Keywords:
Big data analytics
Complexity theory
fsQCA
Business value
Mixed-method
Environmental uncertainty

ABSTRACT

Big data analytics has been widely regarded as a breakthrough technological development in academic and business communities. Despite the growing number of firms that are launching big data initiatives, there is still limited understanding on how firms translate the potential of such technologies into business value. The literature argues that to leverage big data analytics and realize performance gains, firms must develop strong big data analytics capabilities. Nevertheless, most studies operate under the assumption that there is limited heterogeneity in the way firms build their big data analytics capabilities and that related resources are of similar importance regardless of context. This paper draws on complexity theory and investigates the configurations of resources and contextual factors that lead to performance gains from big data analytics investments. Our empirical investigation followed a mixed methods approach using survey data from 175 chief information officers and IT managers working in Greek firms, and three case studies to show that depending on the context, big data analytics resources differ in significance when considering performance gains. Applying a fuzzy-set qualitative comparative analysis (fsQCA) method on the quantitative data, we show that there are four different patterns of elements surrounding big data analytics that lead to high performance. Outcomes of the three case studies highlight the inter-relationships between these elements and outline challenges that organizations face when orchestrating big data analytics resources.

1. Introduction

We are living in the “Age of Data”, with new data being produced from all industries and public bodies at an unprecedented, and constantly growing rate (McAfee, Brynjolfsson, & Davenport, 2012). As a result, there has been a great hype which has led organizations to make substantial investments in their quest to explore how they can use their data to create value (Constantiou & Kallinikos, 2015). The main premise big data analytics builds on is that by analyzing large volumes of unstructured data from multiple sources, actionable insights can be generated that can help firms transform their business and gain an edge over their competition (Chen, Chiang, & Storey, 2012). Being able to obtain such data-generated insight are particularly relevant, especially for organizations that operate in dynamic and high-paced business environments, where making informed decisions is critical (Wamba et al., 2017). Despite much promise from big data analytics, there has been significantly less research on how organizations need to be structured in order to generate business value from such investments, and a limited understanding on the interplay of factors that drive performance gains (Vidgen, Shaw, & Grant, 2017). Most reports to date on the value of big data analytics come from consultancy firms, popular press, and isolated case studies, which fail to build on empirical results from large-scale analyses and lack theoretical insight (Gupta & George, 2016). Furthermore, recent studies have noted that there is still a sizeable number of companies that fail to capture value from their big data investments (Popovič, Hackney, Tassabehji, & Castelli, 2018; Wamba et al., 2017), and even some that argue that big data may hurt rather than help companies (Kiron, 2017). As a result, there is insufficient understanding about how organizations should approach their big data initiatives, and scarce empirical support to guide value creation from such investments (Mikalef, Pappas, Krogstie, & Giannakos, 2018).

Recognizing these issues that many organizations face, several research commentaries have been written that underscore the importance of delving into the whole spectrum of aspects that surround big data analytics (Constantiou & Kallinikos, 2015; Sharma, Mithas, & Kankanhalli, 2014). Nevertheless, empirical studies on the topic are still quite scarce, especially in explaining how performance gains can be

* Corresponding author.
E-mail address: patrick.mikalef@ntnu.no (P. Mikalef).

https://doi.org/10.1016/j.jbusres.2019.01.044
Received 18 July 2018; Received in revised form 22 January 2019; Accepted 23 January 2019
Available online 15 February 2019

achieved, and what factors contribute to their attainment (Mikalef, Framnes, Danielsen, Krogstie, & Olsen, 2017; Vidgen et al., 2017). An emerging body of literature builds on the notion of big data analytics capability, a key organizational capability in effectively leveraging big data analytics resources towards specific business objectives (Gupta & George, 2016). A big data analytics capability is defined as the ability of a firm to capture and analyze data towards the generation of insights by effectively orchestrating and deploying its data, technology, and talent (Mikalef et al., 2018). As such, firms must acquire and develop a combination of data, technological, human, and organizational resources to create a capability that is difficult to imitate and transfer (Vidgen et al., 2017). While there is a growing body of work on defining the resources that are critical in developing a big data analytics capability, the vast majority of studies work under the assumption that there is an absence of heterogeneity in how organizations develop these. In addition, empirical work in this direction builds under the premise that big data analytics resources are of equal importance, regardless of context (Gupta & George, 2016). Understanding the core big data analytics resources upon which firms realize differential value is of increased importance as more and more companies invest heavily in such technologies and delve into data-driven decision-making (Abbasi, Sarker, & Chiang, 2016; Pappas, Mikalef, Giannakos, Krogstie, & Lekakos, 2018).

Grounded on past research which argues that deriving value from big data analytics requires the orchestration of complementary organizational resources, this study posits that depending on the context of examination some big data analytics resources will have a greater or lesser significance in performance gains (Gupta & George, 2016). In doing so, we adopt a complexity theory approach and suggest that organizations will develop different approaches to leverage their big data analytics resources towards the attainment of organizational goals (Woodside, 2014). We build on a sample of 175 survey responses from IT managers in Greek firms, and we examine the patterns of big data analytics resources that lead to high levels of performance. We apply a configurational approach through the novel methodological tool fsQCA, which allows the examination of such complex phenomena and the reduction of solutions to a core set of elements. In addition, we examine three case studies to uncover how these elements, as well as other core enablers or inhibitors emerge, and how they coalesce and impact performance. The overall research questions that guide this research are:

What combinations of big data analytics resources enable firms to achieve high performance and in what contexts?

What critical aspects require consideration when leveraging big data analytics resources?

The rest of the paper is structured as follows. In Section 2 we provide an overview of the literature on big data analytics and business value summarizing the current state of knowledge and highlight the gaps that exist that this study attempts to answer. We then introduce the research framework and outline the theoretical perspective that this study builds on. Section 3 delineates the overall research approach, describes the method of the study, including the data, measurements, reliability and validity tests, as well as the method for gathering data in the case studies. In Section 4 we present the results of the fsQCA analysis and the outcomes of the three case studies. Finally, in Section 5 the findings are discussed, the theoretical and practical implications are highlighted, and limitations and future research directions are presented.

2. Background and research framework

2.1. Big data analytics and business value

Big data analytics have been considered by many as the next frontier for innovation, competition, and productivity (Manyika et al., 2011). As a result, there has been considerable attention from both academics and practitioners on the value that organizations can derive from the use of big data analytics towards the attainment of organizational goals. A widely used definition of big data analytics regards them as "*a new generation of technologies and architectures*, *designed to economically extract value from very large volumes of a wide variety of data, by enabling high velocity capture*, *discovery and/or analysis*" (Mikalef, Pappas, Krogstie, & Giannakos, 2017). The literature suggests that through focused deployment of big data analytics, firms are able to sense emerging opportunities and threats, generate critical insight, and adapt their operations based on trends observed in the competitive environment (Chen et al., 2012). As a result, the major competitive differentiator that big data analytics provides lies in the fact that it facilitates better informed decision-making (Abbasi et al., 2016; Mikalef, Boura, Lekakos, & Krogstie, 2019). The increased interest in big data analytics has been particularly evident in companies operating in complex and fast-pasted environments (Wang, Gunasekaran, Ngai, & Papadopoulos, 2016). Managers nowadays are basing their decisions more and more on real-time insight generated from big data, and are directing a growing number of initiatives in this direction (Constantiou & Kallinikos, 2015). Several research papers demonstrate that big data analytics, when applied to problems of specific domains such as healthcare, service provision, supply chain management, and marketing, can offer substantial value (Mikalef et al., 2019; Raghupathi & Raghupathi, 2014; Waller & Fawcett, 2013; Wang et al., 2016). Adding to these studies, a report by the MIT Sloan Management Review shows that big data analytics can also be a source of innovation, with those companies that are leaders in adoption being more likely to deliver new products and services in comparison to the laggards (Ransbotham & Kiron, 2017).

Nevertheless, despite the many claims that big data analytics can lead to business value, there is still limited knowledge on the organizational aspects and challenges that are important when attempting to do so (Gupta & George, 2016; Wamba et al., 2017). Sharma et al. (2014) highlight that while there is some evidence suggesting that big data analytics can create business value, the thesis that big data analytics leads to business value requires a deeper analysis. A new stream of studies argues that gaining value from big data analytics is a result of focused organizational diffusion of these technologies into operations, and therefore requires a firm-wide big data analytics capability to be developed (Gupta & George, 2016; Mikalef et al., 2019; Wamba et al., 2017). Building on this stream of research and synthesizing definitions, Mikalef et al. (2018) frame a big data analytics capability as the ability of a firm to effectively deploy technology and talent to capture, store and analyze data, towards the generation of insight. In their empirical study Vidgen et al., (2017) note that organizations face several challenges when attempting to generate value out of their big data analytics, and that these challenges have to do with how big data analytics are orchestrated and leveraged. Several business reports seem to point to the same underlying issue in light of the big data analytics phenomenon, with most challenges faced by companies in deriving business value being of an organizational nature (Kiron, 2017). While this issue is becoming increasingly more apparent in academic literature and practice, there is still limited understanding on what organizational aspects are important when attempting to gain business value from big data analytics investments (Abbasi et al., 2016; Mikalef, Framnes, et al., 2017). Even more, there is limited knowledge on how the context influences such capabilities and shapes the factors that are critical in realizing performance gains (Günther, Mehrizi, Huysman, & Feldberg, 2017).

From the empirical work performed to date, there have been several studies that isolate factors that contribute to successful organizational diffusion of big data analytics. For instance, Gupta and George (2016) develop a measure of a firms capability to orchestrate big data analytics and generate performance gains, which distinguishes between tangible resources (e.g. infrastructure, data and financial resources), human skills (technical and business analytics skills), as well as intangible resources (including a data-driven culture and a propensity for organizational learning). Their findings demonstrate that investing in these resources is associated with increased market and operation performance. Similarly, Wamba et al. (2017) empirically showcase how investing in infrastructure, management capabilities and personnel expertise capabilities can lead to gains in overall firm performance. While there is a growing body of research that identifies core areas that contribute to the development of a big data analytics capability (Mikalef et al., 2018), there is an underlying assumption that all firms must focus to an equal extent on these elements. As a result, there is limited heterogeneity in the ways in which firms are suggested to gain value from their big data investments, and rarely is the role of the context included in such investigations. Past research in the broader area of information systems has shown that success of IT projects largely depends on the context in which they are deployed and on several contingency elements (Bechor, Neumann, Zviran, & Glezer, 2010). The main premise on which these studies build on is that depending on the context of examination, there are resources that will have a greater or a lesser importance in realizing performance gains (Petter, DeLone, & McLean, 2013).

Nevertheless, in the case of big data analytics there is still limited research looking into how resource importance may differ based on the context of examination, and how the blend of resources and context may lead to improvements in performance. While research to date has begun to elucidate the role that different elements within these categories have on realizing performance gains from big data analytics investments, there are still limited studies that examine the confluence of contextual factors (Mikalef, Framnes, et al., 2017). Furthermore, most of the studies conducted so far build on the assumption that all organizations face the same challenges, and thus should focus their investments in a uniform set of aspects.

2.2. Research framework

Recent conceptual and empirical research recognizes that the challenge of deriving business value from big data analytics is not solely a technical one, but mostly an organizational one (Gupta & George, 2016). Vidgen et al. (2017) show through a Delphi study and three case studies, that the five main challenges organizations face in becoming data-driven revolve around data, technology, processes, people, and organization. Their empirical work builds upon the five challenges set forth by McAfee et al. (2012), and traces back to work of socio-technical systems and the diamond model of Leavitt (1965). The rationale behind this perspective is that a big data analytics capability is responsible for converting the data that a firm collects into business value by leveraging it into actionable insight. Similar approaches have been described in several other research studies. For instance, Wamba et al. (2017) apply a socio-materialistic perspective in developing their notion of big data analytics capabilities, and identify factors through a resource-based view (RBV) framework. Their work highlights the importance of complementary organizational factors in deriving value from big data analytics and exemplifies their logic by conducting a quantitative study linking big data analytics capabilities with firm performance. Following a similar approach which builds on the RBV, Gupta and George (2016) demonstrate that value from big data analytics is a result of maturity in tangible, intangible, and human-related big data resources. In their paper, the authors argue that amongst others, intangible resources such as a data-driven culture and a propensity towards organizational learning are key components in driving business value. Building on the same set of resources that form a big data analytics capability, Mikalef et al. (2019) showcase the mechanisms through which value can be derived, and find positive effects on a firms incremental and radical innovative capabilities. In a recent systematic literature review, Mikalef et al. (2018) provide a conceptual distinction between big data analytics, and big data analytics capabilities, highlighting the importance of the latter when considering performance gains. The authors overview studies on business value of big data analytics and present a research framework that highlights the importance of factors that pertain to processes, people, technology, organization and data. In consonance with academic literature, business reports also indicate that the biggest challenges managers face when attempting to derive value from big data are organizational ones (Kiron, 2017; LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011).

While these studies and reports outline complementary factors of big data analytics that help realize business value, they fail to examine their interdependencies as well as their significance under different contextual conditions. Several research commentaries have argued that it is important to examine under what circumstances big data analytics capabilities add value, particularly due to the high costs associated with developing them (Abbasi et al., 2016). In addition, the resource and decision-making structures that govern their interdependencies is an area that to date has received inadequate attention (Sharma et al., 2014). An emerging theme in big data analytics and business value research is that companies differ in the way they operate, and thus require attention in different sets of aspects. Findings from business reports highlight that many laggard companies attempt to imitate forerunners of big data analytics adoption, and frequently fail since they do not take into account the particularities of the context in which they operate (Kiron, 2017). While data, technology, people, processes and the organization of these comprise core components of realizing performance gains, the ways in which they are structured and the extent to which they are important is argued to differ depending on a number of contextual factors. In addition, there is to date very limited research on the relationship between these characteristics and how they affect each other (Mikalef et al., 2018). We therefore propose the research framework presented in Fig. 1.

The Venn diagram illustrates the seven sets of constructs and their intersections. The constructs reflect the outcome of interest of this study which is performance (dependent variable) and six sets of causal categories of variables to predict the outcome (independent variables). The intersections represent factor configurations, which are higher-level interactions. The Venn diagram is a useful way of illustrating the possibilities of the presence and absence of ingredients in complex antecedent conditions (i.e. solutions) indicating high scores in an outcome condition, in this case firm performance (Woodside, 2014). To provide an illustrative case and exemplify the research framework of the Venn diagram, we can develop a hypothetical solution to demonstrate the types of outcomes that are enabled through a complexity theory lens and configurational methodologies. For instance, we can assume that one possible cluster of companies (i.e. solution) that achieve high levels of performance from big data analytics are large firms which operate in highly heterogeneous environments, and do so by investing in data, strong data analytics business skills and establishing solid procedural and structural governance practices. Therefore, high levels of maturity in these specific elements enable the firm to achieve performance gains under a certain set of contextual conditions.

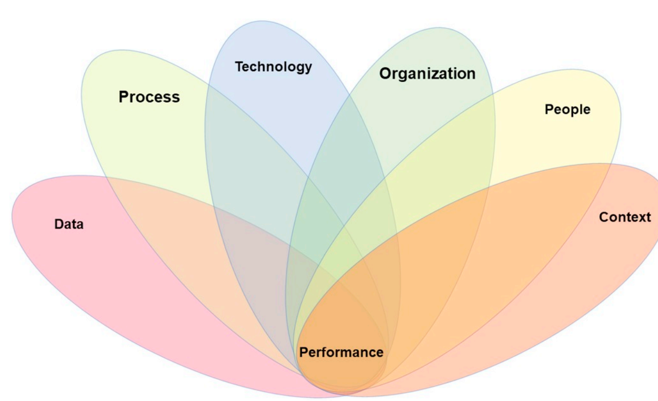

Fig. 1. Research framework.

Nevertheless, one of the main tenants that the research framework, and consequently the research approach and methodology build upon, is that it is possible to uncover multiple different ways through which firms can achieve high performance. In other words, there are multiple realities as Woodside (2014) notes, which highlight different combinations of factors that help produce a specific outcome of interest. We describe the theoretical perspective that supports this view in the following subsection.

2.3. Theoretical perspective

The current research builds on the theoretical groundings of complexity theory, which views organizations as complex adaptive systems that self-organize and evolve to become better suited to their environments (Cabrera, Cabrera, Powers, Solin, & Kushner, 2018). Complexity theory describes several key tenets that differentiate it to variance-based theories. These include the following: (1) there is no single antecedent condition that is sufficient or necessary for describing a high score in an outcome, (2) a few of many available complex configurations of antecedent conditions are sufficient indicators of high scores in an outcome condition, (3) contrarian cases occur, which means that low scores in a single antecedent can be associated with both high and low scores of an outcome condition for different cases, and (4) causal asymmetry can occur, meaning that accurate causal models for high scores of an outcome condition are not the mirror opposites of causal model for low scores for the same outcome condition (Wu, Yeh, & Woodside, 2014). Typically, variance-based approaches offer one single solution, considered as the best solution, that explains the outcome, leaving however a significant amount of the outcome unexplained. Furthermore, focusing on net effects may be misleading (Woodside, 2013a), since besides the main relation amongst the variables, an opposite relationship will exist for some cases in the same sample, thus creating the need to test the data for such contrarian cases (Woodside, 2014). To this end, different configurations of the examined variables may lead to the same outcome depending on how they combine with each other. Such configurations lead to multiple solutions, which in total represent a larger part of the sample and are likely to explain a greater amount of the outcome.

Complexity theory has gained eminence over the past few years in the domains of economics, marketing, psychology, operational, and information systems research (Fiss, 2011; Park, El Sawy, & Fiss, 2017; White, 2018; Wu et al., 2014). The aim of complexity theory is to identify patterns and combinations of conditions and reveal how their synergistic effects lead to specific outcomes (Mikalef & Pateli, 2017). Configurations occur as different combinations of causal variables that affect an outcome of interest (El Sawy, Malhotra, Park, & Pavlou, 2010). The main difference of complexity theory is that it views elements through a holistic lens that must be examined simultaneously, and is therefore particularly attractive for context-related studies looking into complex causality (Woodside, 2013b). Organizational deployment of big data analytics fit well into the lens of complexity theory, since multiple interacting actors, objects, processes and contextual elements shape realized business value (Wilden, Devinney, & Dowling, 2016). In addition the interactions between these components of such complex systems give rise to emergent properties that cannot be fully understood by examining the individual components (Fiss, 2011). Seeing that big data analytics are applied in different ways depending on a number of internal and external organizational factors, applying a complexity theory perspective to examine emergent properties such as performance gains, is deemed as appropriate (Anderson, 1999). A substantial body of literature builds on the theoretical tenants of complexity theory by utilizing the novel methodological approach fsQCA to examine phenomena in organization science (Fiss, 2007, 2011), marketing (Woodside, 2013a), service science (Wu et al., 2014), and information systems research (Mikalef & Pateli, 2017). Researchers have traditionally conducted data analysis and hypothesis testing to examine the symmetric relationship between X and Y. Nevertheless the presence of asymmetrical relationships in most real-life contexts has signalled a theoretical and methodological shift (Woodside, 2013a). Therefore, this study builds on this call as well as on past empirical studies that are grounded in complexity theory and appropriate methodological approaches which are described below.

3. Method

3.1. Research approach

The purpose of this research is to understand how factors relating to a firm's big data analytics resources coalesce with contextual elements to high firm performance. Therefore, we employ a multi-method approach that builds on both quantitative and qualitative methods. The goal of opting for this approach is to (1) identify the combinations of factors that are important in driving performance from big data analytics in different contexts, and (2) uncover the relationships that characterize these configurations. On the one hand, the use of fuzzy-set qualitative comparative analysis (fsQCA) can provide a novel lens of uncovering the different paths of data-driven value and highlight core big data analytics resources in achieving this. On the other hand, through case studies, and specifically though a series of semi-structured interviews with key personnel within firms, we can understand more about the associations that characterize core factors within solutions, and how they emerge to drive performance gains. This allows us to generate deep insight on big data analytics in the organizational setting and provides a basis for drawing a future research agenda.

To explore the research questions we adopted a sequential explanatory strategy, first starting with the quantitative analysis, and then using a qualitative study with semi-structured interviews to gain a better understanding on the meaning and emergence of solutions (Venkatesh, Brown, & Bala, 2013). The quantitative study, which was based on the survey method, builds on factors that have recurrently been noted in literature as being important contributors to big data analytics success (McAfee et al., 2012; Mikalef et al., 2018). Specifically, we build on the dimensions identified in the paper of Vidgen et al. (2017) and follow their call to explore the significance of the elements in driving business value. We utilize the novel fsQCA method to analyze data, since it builds on the main tenets of complexity theory and is well suited for examining situations where an outcome of interest can be achieved in multiple different ways. In addition, there have been a large number of studies on exploring core challenges and factors in driving value from big data analytics, and there is lack of confirmatory research to put these elements to test. Independently from the quantitative study, we conducted three case studies to develop deeper knowledge on how the solutions of the quantitative study emerge. Semi-structured interviews with key respondents in firms represent a powerful way to understand the dynamics that characterize the core constituents of solutions, and why some factors are important under certain conditions and not others. The interview questions were driven by the results of the quantitative study, in combination with an interview guide.

3.2. Survey, administration and data

To examine the significance of factors towards the attainment of performance gains, a survey instrument was developed and administered to key informants within firms. A survey-based approach is deemed as an appropriate method of accurately capturing the maturity of firm's big data analytics capabilities. According to Straub, Boudreau, and Gefen (2004), survey-based research is suitable in exploratory settings and predictive theory. All constructs and their corresponding survey items are based on previously published latent variables with psychometric properties that support their validity. To operationalize concepts and respective constructs, we utilized a 7-point Likert scale, a well-accepted practice in large-scale empirical research where no

Table 1
Descriptive statistics of the sample and respondents.

Factors	Sample (N = 175)	Percentage (%)
Industry		
Bank & financials	19	10.8%
Consumer goods	17	9.7%
Oil & gas	5	2.8%
Industrials (construction & industrial goods)	13	7.4%
ICT and telecommunications	35	20.0%
Technology	16	9.1%
Media	13	7.4%
Transport	3	1.7%
Other (shipping, basic materials, consumer services etc.)	54	30.8%
Firm size (number of employees)		
1–9	34	19.4%
10–49	42	24.0%
50–249	53	30.2%
250+	46	26.2%
Total big data analytics experience		
< 1 year	26	14.8%
1–2 years	38	21.7%
2–3 years	49	28.0%
3–4 years	34	19.4%
4+ years	28	16.0%
Respondent's position		
CEO/president	23	13.1%
CIO	129	73.7%
Head of digital strategy	4	2.0%
Senior vice president	6	3.4%
Director	6	3.4%
Manager	7	4.0%

standard measures exist for quantifying notions such as resources and capabilities (Kumar, Stern, & Anderson, 1993). To validate the statistical properties of the measures and to examine their comprehensiveness, a small-cycle pre-test study with 17 firms was conducted. These firms operated in Greece but were not part of the sample used in the main study. The pre-testing procedure enabled us to assess the face and content validity of items and to ensure that key respondents would be in place to comprehend the survey as intended. These respondents were then contacted by phone and asked about the quality of the questions and invited to provide suggestions in order to improve clarity and presentation of the survey instrument. In response to this feedback some minor modifications were made to the phrasing of the questions.

As part of the main study, a mailing list of approximately 1500 Chief Information Officers and IT managers based in Greece was used. To make sure that all items were answered appropriately, the respondents were instructed to consult other employees within their firms for information that they were not knowledgeable about. Data was collected over a period of approximately three months (April 2017–July 2017), and on average completion time of the survey was 14 min. In total, 193 firms completed the survey, with 175 providing complete responses (Table 1). Firms in our sample operated in various industries, the largest of which was from the ICT sector (20.0%), followed by bank & financials (10.8%), consumer goods (9.7%), technology (9.1%), while a large proportion came from other sectors (30.8%). The majority were medium-sized firms, accounting for 30.2% of the sample, while high percentages were obtained from large-sized (26.2%) and small firms (24.0%). Respondents predominantly occupied senior positions relating to business and IT management as initially projected, while their experience with big data analytics ranged, with most having at least 2 years prior engagement in such technologies.

To examine the possibility of non-response bias in our sample, the profile of the respondents from the mailing list was used to extract information about the industry, type and size-class of the firm. Chi-square analyses on these attributes between responding and non-responding firms, revealed no systematic response bias since there was no significant difference in the number of responding and non-responding firms in terms of the previously mentioned attributes. Furthermore, we examined the possibility of late response bias by comparing early (first two weeks) and late responses (last two weeks) on the main constructs of the study. Outcomes confirmed that there was no statistically significant difference between the two sub-groups. To determine if there is risk of method bias in our sample, we followed the guidelines of Podsakoff, MacKenzie, Lee, and Podsakoff (2003) and conducted a number of analyses. *Ex-ante*, respondents were assured that all information they provided would remain completely anonymous and confidential, and that any analysis would be done on an aggregate level for research purposes solely. *Ex-post*, we run a Harmon one-factor test on the main variables of our study. Outcomes suggest that there is no uni-factor solution since the maximum variance explained by any one factor was 38.1%, an indication of an absence of common method bias. In terms of sample size, the 175 completed responses exceed requirements for latent variables constructs that require ten times the largest number of formative indicators used to measure one construct (Hair, Ringle, & Sarstedt, 2011). Furthermore, unlike conventional statistical techniques, fsQCA overcomes limitations related to sample size (Mas-Verdú, Ribeiro-Soriano, & Roig-Tierno, 2015). Therefore, fsQCA analyses are equally conclusive for small or large samples, making it an appropriate tool for a wide range of research (Fiss, 2011; Navarro, Llinares, & Garzon, 2016; Woodside, 2012).

3.3. Measurements

The scales for the various constructs were adopted from prior literature and have therefore been previously tested in empirical studies. Appendix A provides a summary of the scales used, their descriptive statistics, and the supporting literature. The constructs used are grouped under the dimensions of the conceptual model described in the previous section.

Data (*DATA*) was operationalized as a formative first-order construct. The items of the construct were developed so as to capture the extent to which an organization has access to large, unstructured, and fast-moving data and the degree to which it integrates its internal data and external data (Gupta & George, 2016).

Technology (*TECH*) was developed as a formative first-order construct consisting of five items. The construct identifies whether an organization possesses the necessary data storage technologies, data visualization tools, and other complementary cloud-based and open-source data analytics infrastructure (Gupta & George, 2016; Vidgen et al., 2017).

People were assessed based on two constructs pertaining to the respective type of skills, *Technical Skills* (*TSKL*) and *Managerial Skills* (*MSKL*). Both notions were operationalized as first-order reflective constructs and measured the degree to which technical and managerial staff had big data analytics-specific skills. Technical skills assessed the level to which staff had the right skills to accomplish their jobs successfully, as well as if there is suitability in training and education background in relation to big data analytics requirements (Gupta & George, 2016). Managerial skills on the other hand examined the degree to which managers were knowledgeable about areas to apply big data analytics, understood the business needs of different functional areas and the opportunities big data analytics allow for, as well as their knowledge on how to evaluate the output extracted from big data.

Organization of big data analytics within firm operations was gauged by looking at the extent to which companies have established *Structural* (*STRU*) and *Relational* (*RELA*) practices, as well as how effective they have been in infusing a strong *Data-driven Culture* (*CULT*) within firm boundaries. Structural and relational practices are part of a firms big data, or information, governance schemes (Tallon, Ramirez, & Short, 2013). Structural practices define key IT and non-IT decision makers and their corresponding roles and responsibilities when it

Table 2
Assessment of reliability, convergent and discriminant validity of reflective constructs.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Data	n/a											
(2) Technology	0.24	n/a										
(3) Technical skills	0.52	0.26	0.88									
(4) Managerial skills	0.56	0.31	0.37	0.91								
(5) Structural practices	0.47	0.48	0.30	0.57	0.90							
(6) Relational practices	0.27	0.32	0.22	0.30	0.34	0.93						
(7) Data-driven culture	0.53	0.36	0.38	0.51	0.37	0.35	0.88					
(8) Procedural practices	0.06	0.17	0.16	0.23	0.14	0.23	0.18	0.82				
(9) Performance	0.12	0.14	0.19	0.21	0.21	0.15	0.26	0.43	0.85			
(10) Dynamism	0.17	0.10	0.20	0.09	0.19	0.26	0.19	0.31	0.24	0.93		
(11) Heterogeneity	0.23	0.26	0.11	0.28	0.32	0.26	0.25	0.37	0.27	0.31	0.93	
(12) Hostility	0.41	0.37	0.43	0.26	0.28	0.31	0.31	0.18	0.33	0.16	0.26	0.94
Mean	4.98	4.61	4.51	5.07	4.45	4.10	5.01	5.03	3.94	4.67	4.13	4.79
Standard deviation	1.72	2.02	1.82	1.84	1.95	1.51	1.81	1.82	1.39	1.45	1.34	1.64
AVE	n/a	n/a	0.77	0.82	0.81	0.86	0.75	0.68	0.73	0.87	0.86	0.89
Cronbach's alpha	n/a	n/a	0.90	0.93	0.76	0.84	0.83	0.88	0.79	0.91	0.90	0.86
Composite reliability	n/a	n/a	0.93	0.95	0.89	0.92	0.90	0.91	0.88	0.92	0.91	0.89

comes to data ownership, value analysis, and cost management. Structural practices include, explicit declarations about the main roles of setting policies and standards for protecting and using data. Relational practices on the other hand are concerned with the formalized links between employees of the technical and business sides. They encapsulate practices and means of knowledge sharing, education and training, and strategic planning (Kooper, Maes, & Lindgreen, 2011). Data-driven culture assesses whether an organization considers its data a tangible asset, and determines the extent to which organizational decisions are made based on the extracted insight (Gupta & George, 2016). In essence, the construct captures the importance firms give on data, and the extent to which they base their decisions on insight rather than instinct. All constructs are developed as first-order reflective latent variables, with several underlying items.

Processes regarding big data analytics refer to the formal methods for managing and leveraging data in order to generate insight. In this regard, *Procedural* (*PROC*) practices – typically part of big data governance – are concerned with activities that amongst others include data migration, data retention, cost allocation, data analytic procedures, and access rights. These organizational practices can differ based on the type of data analyzed, or the type of insight that is explored (Mikalef & Pateli, 2017). Procedural practices were operationalized as a first-order reflective construct (Tallon, 2013).

Performance (*PERF*) was operationalized by measuring profitability, market share, growth, innovativeness, cost leadership, and delivery cycle time in relation to main competitors (Liu, Ke, Wei, & Hua, 2013; Rai & Tang, 2010). These measures are representative of the potential value that can be realized as a result of strong big data analytics capabilities (Gupta & George, 2016; Vidgen et al., 2017; Wamba et al., 2017). Performance was developed as a first-order reflective construct consisting of 10 indicators and represents the dependent variable of this study.

Contextual variables. Firm size was measured as an ordinal value in accordance with the recommendations of the European Commission (2003/361/EC) into micro (0–9 employees), small (10–49 employees), medium (50–249 employees), and large (> 250 employees). The uncertainty of the external environment was defined as the degree of unpredictability and imperfect knowledge about the environment (Verdu, Tamayo, & Ruiz-Moreno, 2012). The notion was developed through three first-order reflective constructs, with dynamism (DYN), heterogeneity (HET), and hostility (HOST) (Newkirk & Lederer, 2006). Dynamism reflects the rate and unpredictability of environmental change, heterogeneity the complexity and diversity of external factors, and hostility as the availability of key resources and the level of competition in the external environment.

3.4. Measurement model

To establish that the used constructs were valid and reliable measures we performed a series of analyses through the software package SmartPLS 3. (Ringle, Wende, & Becker, 2015). Since the research design contains both reflective and formative constructs, as well as higher-order variables, we used different assessment criteria for each. First-order reflective latent constructs were subjected to reliability, convergent validity, and discriminant validity tests. Reliability was examined at the construct and item level. At the construct level we examined the Composite Reliability (CR), and Cronbach Alpha (CA) values, and confirmed that their values were above the threshold of 0.70 (Nunnally, 1978). At the indicator level, we examined if construct-to-item loadings exceed the lower limit of 0.70, with all values surpassing this threshold. To examine if convergent validity was met, we looked at AVE values to verify that they exceeded the lower limit of 0.50, with the smallest value observed being 0.68. Next, we verified discriminant validity in three ways. The first looked at each constructs AVE square root to determine if it was greater that its highest correlation with any other construct (Fornell-Larcker criterion). The second examined if each indicator's outer loading was greater that its cross-loading with the other constructs used in the study (Farrell, 2010). Third, we employed the heterotrait-monotrait ratio (HTMT), a criterion argued by Henseler, Ringle, and Sarstedt (2015) as being a better assessment indicator of discriminant validity. All values where above the lower threshold of 0.85, a strong indication of discriminant validity. The previously mentioned outcomes, as presented in Table 2 and Appendix B suggests that first-order reflective measures are valid to employ in further analyses, and that all items are good indicators of their respective constructs.

To assess the appropriateness of formative indicators, we first examined the weights and significance of items to their assigned construct. The items of all first-order constructs had positive and highly significant effects. To examine the validity of the items of formative constructs, we followed the guidelines suggested by MacKenzie, Podsakoff, and Podsakoff (2011) and Vom Brocke et al. (2014) and calculated Edwards (2001) adequacy coefficient (R_a^2). To do so, we summed the squared correlations between formative items and their respective formative construct and then divided the sum by the number of indicators. All R_a^2 values surpassed the threshold of 0.50 (Table 3), indicating that the majority of variance in the indicators is shared with the overarching construct, and that the indicators are valid representations of the construct (Edwards, 2001). Finally, we also looked at the extent to which the indicators of formative constructs presented multicollinearity. Variance Inflation Factor (VIF) values below 10 suggest low multicollinearity, however, a more restrictive cut-off of 3.3

Table 3
Formative construct validation.

Construct	Measures	Weight	Significance	VIF	R_a^2
Data	D1	0.383	$p < 0.001$	2.800	0.79
	D2	0.287	$p < 0.001$	1.300	
	D3	0.552	$p < 0.001$	1.112	
Technology	T1	0.209	$p < 0.001$	2.256	0.76
	T2	0.398	$p < 0.001$	1.986	
	T3	0.358	$p < 0.001$	2.285	
	T4	0.202	$p < 0.001$	2.129	
	T5	0.552	$p < 0.001$	2.030	

is used for formative constructs (Petter, Straub, & Rai, 2007). All values were below the threshold of 3.3 indicating an absence of mutlicollinearity.

3.5. Qualitative data

To explore how factors under the five main categories coalesce and drive business value in different contextual conditions, we conducted three case studies in private organizations. The choice of a case study design is most appropriate when there is a need to understand how phenomena emerge in a specific context (Yin, 2009). In addition, we chose this approach as we wanted to observe the phenomenon of how big data analytics capabilities emerge in real business settings, as well as how they differ depending on a number of internal and external conditions. Using multiple case studies allowed us to identify technical, organizational, and contextual aspects related to implementation, as well as to study the interactions that develop between these components. Thus, the aim of the case studies was to complement the quantitative analysis and to provide more depth on the solutions that emerge. Furthermore, the case studies were used as a basis to uncover other aspects that can potentially enable or inhibit performance gains that were not included in the quantitative study.

Data were collected over a period of two months through semi-structured interviews with the chief digital officer, chief information officer, IT manager or project manager of each organization. Although interviews are a highly efficient way to collect rich data, there is the limitation that information gathered is rather subjective, since it originates from respondents within the firm. We mitigated this threat by collecting data from primary and secondary data sources for each firm. The primary source of data was direct interviews with the key respondents, during which their attitudes, beliefs, and opinions were asked regarding their experience with big data analytics initiatives their firm had undertaken. The interview guidelines and the respective questions can be found in Appendix D. Interviews were performed face-to-face in a conversational style, opening with a discussion on the nature of the business and then proceeding on to the themes of the interview guideline. When necessary, questions were clarified to encourage more accurate responses. Discussions were recorded and later transcribed for analysis. Each interview lasted between 65 and 80 min. To corroborate statements of the interviewees, secondary data sources were utilized, including published information about the firms' annual reports, presentations, meeting notes, website information, as well as third-party news articles.

The empirical analysis was performed by an iterative process of reading, coding, and interpreting the transcribed interviews and observation notes of the three case studies (Myers & Newman, 2007). During the first phase, we identified and isolated a large number of concepts on the grounds of the theoretical distinction we developed in the quantitative study. For each case the standardization method was used to quantify these characteristics using an open coding scheme (Yin, 2017). This allowed us to cluster primary data in a tabular structure, and through an iterative process identify the relative concepts and notions that were applicable for each case. Two of the co-authors completed the independent coding of the transcripts in accordance with the defined themes. Each coder read the transcripts independently to find specific factors related to the core resources of a big data analytics capability, as well as on business value derived from such their combinations. This process was repeated until inter-rater reliability of the two coders was > 90% (Boudreau, Gefen, & Straub, 2001).

The case studies were selected on the grounds that they were involved in big data analytics projects for at least one year, and that they had established a dedicated group in this area. All cases are reported pseudonymously to retain anonymity and include the following:

Avicom is a state-owned company that operates 45 airports in Europe and also operates air traffic control towers, control centers, and technical infrastructure for aircraft navigation serving the civil and military aviation. In addition, it is involved in letting space for duty-free shops, cafés, and restaurants, as well as space for airport hotels and parking facilities. Avicom was founded in 2003, and as of 2018 employs > 2400 people. To gather information from Avicom three employees were interviewed, the head of business intelligence, a project management, and a data analyst. Each had substantial experience in the company having worked there for over 8 years.

GLM is one of the four major banks in Greece with > 7 million customers, 500 branches in Greece and abroad, and a market share of over 25% in the Greek market. It provides financial products and services to corporate and retail customers. The bank must deal with the ongoing economic crisis in Greece that has significantly affected its profits due to non-performing mortgages, loans and credit cards debts. Two people were interviewed in GLM, the chief information officer and an IT manager whom were both involved since the beginning in the planning and implementation of big data analytics projects.

DataCom is a newly found business operating in Greece that offers services in the areas of social media analytics. DataCom collects and analyzes real-time data coming from social media and the web to extract and deliver hidden knowledge from user generated content. Social media and web data have evolved to a valuable source of customer insights, which when properly mined can provide significant value for the firm and its clients. To collect information regarding the activities of DataCom in big data analytics, the chief information officer and a senior project manager were interviewed. Both were founding members of the firm and were highly knowledgeable about activities, investments and events that had occurred during the adoption of big data analytics.

These cases were analyzed individually before we conducted a cross-case analysis.

4. Findings

To determine what combinations of big data analytics resources are most important in the attainment of performance for firms operating in varying contexts, this study employs a fuzzy-set Qualitative Comparative Analysis (fsQCA). FsQCA follows the principles of complexity theories in a configurational approach which allow for the examination of interplays that develop between elements of a messy and non-linear nature (Fiss, 2011). The main difference of fsQCA with other statistical methods is that it supports equafinality, meaning that a particular outcome (e.g. high levels of firm performance) may be caused by different combination of elements, and that these combinations of elements may differ depending on context. This is particularly relevant to the case of big data analytics since depending on the areas towards which insight generation is targeted, the factors that are core contributors to firm performance may vary significantly (Abbasi et al., 2016). As such, it is important to isolate the combinations of factors and conditions that enable firms to achieve high performance outcomes. FsQCA follows such a paradigm since it is geared towards reducing elements for each pattern to the fundamentally necessary and sufficient conditions. In addition, fsQCA further supports the occurrence of causal asymmetry. Causal asymmetry means that, for an outcome to occur, the presence and absence of a causal condition depend on how this causal condition combines with one or more other causal conditions (Fiss, 2011).

4.1. Calibration

The first step of the fsQCA analysis is to calibrate dependent and independent variables into fuzzy or crisp sets. Performance is set as the dependent variable of our study, while the independent variables that are used include data, technology, technical and managerial skills, structural, relational and procedural practices, data-driven culture, as well as elements of the external environment such as dynamism, heterogeneity, hostility, and the size-class which firms belongs to. Fuzzy sets may range anywhere on the continuous scale from 0, which denotes an absence of set membership, to 1, which indicates full set membership. Crisp sets are more appropriate in categorical variables that have two, and only two options such as a firm's size-class which is dichotomized into large firms with 250 or more employees and Small-Medium Enterprises (SMEs) with < 250 employees. Fuzzy sets on the other hand are best suited in converting continuous values such as all other constructs that are on a 7-point likert scale. To calibrate continuous variables into fuzzy sets we followed the method proposed by Ragin (2009). According to the procedure, the degree of set membership is based on three anchor values. These represent a full set membership threshold value (fuzzy score = 0.95), a full non-membership value (fuzzy score = 0.05), and the crossover point (fuzzy score = 0.50) (Woodside, 2013b). Since this study uses a 7-point Likert scale to measure constructs, the suggestions put forth by Ordanini, Parasuraman, and Rubera (2014) are followed to calibrate them into fuzzy sets. Following these guidelines, and based on prior empirical research (Fiss, 2011; Ragin, 2009), we computed percentiles so that the upper 25 percentiles serve as the threshold for full membership; the lower 25 percentiles for full non-membership; and the 50 percentiles represent the cross-over point. Appendix C shows the thresholds for the variables included in this study and the anchor values for each.

4.2. Fuzzy set qualitative comparative analysis

For the analysis of configurations leading to high performance outcomes we relied on the software fsQCA 3.0 (Ragin, 2009). By applying the fsQCA algorithm a truth table of 2^k rows is produced, where k is the number of predictor elements, and each row indicates a possible combination. FsQCA then sorts all the 175 observations into each of these rows based on their degree of membership of all the causal conditions. Consequently, some truth table rows may contain many cases and others just a few or even none. At this stage it is necessary to reduce the number of rows according to two conditions: (1) a row must contain a minimum number of cases, this value was set to a frequency threshold of 5 cases (Ragin, 2009); and (2) selected rows must achieve a minimum consistency level of 0.80. Consistency measures the degree to which a subset relation has been approximated. It resembles the notion of significance in statistical models (Schneider & Wagemann, 2010). Thus, solutions that do not adhere to this threshold are not included in the analysis. Solution coverage on the other hand assesses the empirical relevance of a consistent subset, an analogous measure of R^2 in regression analysis (Mendel & Korjani, 2012). Overall, 15 potential configurations/rows fulfilled these conditions, and these included a total of 138 observations.

After running the fsQCA analysis, an algorithm based on Boolean algebra, the truth table rows are logically reduced to simplified configurations of causal conditions that are necessary to yield high performance outcomes. The fsQCA analysis yields two types of solutions, the intermediate solution (includes simplifying assumptions based on easy counterfactuals) and the parsimonious solution (includes all simplifying assumptions regardless of whether they are based on easy or difficult counterfactuals) (Ragin & Fiss, 2008). To obtain results we use

Table 4
Configurations of big data analytics resources that lead to high firm performance.

Configuration	Solution			
	High firm performance			
	1	2	3	4
Data	●	●	●	●
Technology	●	●		
People				
Technical skills	●	●	●	●
Managerial skills	⊗	●	●	●
Organization				
Structural practices			●	●
Relational practices				●
Data-driven culture			●	
Process				
Procedural practices	●		●	●
Context				
Dynamism	●	●	●	●
Heterogeneity	⊗	⊗		●
Hostility			●	
Large Firms	●		●	
Small-Medium Enterprises (SME's)		●		●
Consistency	0.893	0.957	0.911	0.876
Raw coverage	0.258	0.180	0.253	0.192
Unique coverage	0.232	0.132	0.182	0.163
Overall solution consistency	0.845			
Overall solution coverage	0.523			

the method proposed by Ragin and Fiss (2008) by identifying core conditions that are part of both parsimonious and intermediate solutions, and peripheral conditions are those that are eliminated in the parsimonious solution and only appear in the intermediate solution (Fiss, 2011). Outcomes of the fuzzy set analysis for high levels of firm performance are presented in Table 4. The black circles (●) denote the presence of a condition, while the crossed-out circles (⊗) indicate the absence of it (Ragin, 2008). Core elements of a configuration are marked with large circles (prime implicants which are produced by the parsimonious and intermediate solution of fsQCA), peripheral elements with small ones (implicants that are present in intermediate solutions but not in the parsimonious solutions), and blank spaces are an indication of a don't care situation in which the causal condition may be either present or absent. In the solutions of the present study no peripheral elements exist.

In Table 4, each column represents an alternative combination of conditions that associate to the respective outcome; in this case, high firm performance. The analysis reveals that there are four alternative solutions that lead to high performance. Each solution represents a cluster of firms that share common configurations of elements, or antecedents, that are linked to high levels of performance. The first two solutions present some commonalities since they refer to firms that operate in contexts of high dynamism and an absence of heterogeneity which include companies in industries such as consumer goods, media, transport, and industrials. Such industries present low complexity since the products and services offered don't typically span multiple domains and are characterized by frequent changes in customer requirements and fierce competition. Solution 1 corresponds to large firms, while solution 2 to Small-Medium Enterprises (SME's). For both solutions, data and supporting technological resources represent core aspects in realizing performance gains. In addition, strong technical skills are marked as necessary in converting data into actionable insight. The only observed difference concerns the need for establishing procedural practices in large firms, which is found to be a non-important element in SME's. These findings highlight that in many cases, due to the scale and type of data analytics projects, SME's do not prioritize the formulation of well-defined practices about procedures surrounding data management. In some cases where the scale and complexity of projects, and volume and variety of data may be limited, such processes may also be a cause of rigidity rather than an enabler of value (Chen & Zhang, 2014).

Solutions 3 and 4 correspond to conditions of higher uncertainty. Specifically, solution 3 presents core resources for large firms that operate under conditions of high dynamism and hostility. Such industries may include for instance oil & gas, as well as those in the banking and financial sector. In solution 3 the importance of technology is lesser compared to that in solutions 1 and 2, as it is not found to be a core element of performance attainment. This can be justified by the fact that technology may not be a differentiating element of performance but rather a commodity. Data and technical skills continue to be important factors. In addition, managerial skills emerge as a core element, demonstrating that under highly uncertain market conditions, managerial bandwidth capable of solving business problems through analytics is a core contributor to high performance. Furthermore, under such conditions strong structural and procedural practices, as well as a firm-wide data-driven culture are critical components. These factors underscore the importance of a clear strategy with regards to big data analytics, since these initiatives are most commonly driven by top management. Moreover, they represent a greater fusion with strategic directions and a move towards a more data-driven decision-making structure. The results of solution 4 indicate that for SME's, operating in conditions of high dynamism and heterogeneity, a similar pattern of core resources emerges. Again, managerial skills are critical, as are data resources and technical skills. Establishing structural, relational and procedural practices are also found to be critical components of high performance, denoting that more detailed big data governance schemes need to be established as uncertainty increases and big data analytics becomes a core part of operations.

The findings provide support for the idea that different combinations of big data analytics resources play a greater or lesser importance depending on the contexts of application and the conditions that characterize them. Our results show that different combinations of resources are found to be significant contributors to firm performance depending on characteristics of the external environment as well as on the size-class of the focal firm. Results point out that there exists equifinality in value-creating configurations and also hint that the relationship between maturity of resources and firm performance is not always linear.

4.3. Case studies

To further explore how big data analytics is leveraged towards value creation is different contexts, we investigated three organizations. These organizations presented different characteristics, and therefore aligned well with our attempt to exemplify equifinality in achieving business value through big data analytics. This approach was deemed as most suitable, since it allows us to understand how the factors examined in the quantitative study coalesce, and the dynamics that evolve between them in realizing performance gains (Venkatesh et al., 2013). Thus, the goal of the case studies was to further explore the interdependencies of core big data analytic capability resources and uncover emerging themes. These elements that emerged through the case studies were grouped into: (1) big data analytics strategy, (2) organizational inertia, and (3) ethics and legislation.

4.3.1. Big data analytics strategy

All three cases noted that big data analytics strategy was a significant contributor to attaining performance gains. Several aspects of big data analytics strategy were mentioned including having a clear roadmap for the future, developing a top-down strategy, and having a sense of direction about how analytics can improve business. Respondents highlighted that big data analytics strategy was something that developed gradually in their organizations. Initial experimentation with big data analytics was performed within the IT department, with some early successful business cases leading to a gradual understanding of the importance of analytics, and a more sophisticated view of how it can be linked to strategy. In all cases big data analytics strategy was not so clear but evolved gradually after experimenting with data and demonstrating business value. On the one hand using such a bottom up approach allows for a gradual maturing of big data analytics capabilities, on the other hand however it can severely hamper potential due to a lack of resources in the early stages of experimentation.

The respondents highlighted concerns about the lack of a top-down strategy and the limitations it had in realizing business value. Specifically, Avicom noted that getting management to invest more resources and connect analytics to strategy has been a continuous process:

> "*We started out with aiming for some low-hanging fruit…showing to the business side that data-driven decision has value and that these are some areas that they can cut costs on. Now we have more data than we can process…what we are missing are some people with experience in analytics…but we have to make a strong case why we need them.*"
>
> (*Avicom*)

> "*We must admit that there very few real* "*expert*" *data scientists in the market. Sometimes people tend to call themselves data scientists because they just know how to apply basic statistical techniques or they are using excel.*"
>
> (*GLM*)

The lack of resources due to an unclear strategy surrounding big data analytics is one of the aspects that respondents mentioned. The other concerns the loose connection to the strategic direction of the organization. Big data analytics is seldom the driver of business strategy and is mostly used to improve operational inefficiencies. This was stated by the large companies in our case study, where strong management structures were more prevalent. In these cases, respondents stressed that analytics is not discussed much in top management strategy formulation. In Avicom and GLM these structures constituted a barrier in placing analytics as an enabler of strategy, while in DataCom the relatively small firm size and the flat structure has allowed it to drive its strategy based on analytics. This denotes that there are path dependencies that act as rigidities when considering the fusion of big data analytics into corporate strategy. AVICOM has recognized this issue and is planning to move towards a structure where there is a dedicated analytics manager sitting in the board of advisors:

> "*It is not easy with the current organization to move the company towards more analytics-driven…we are planning to re-structure this…the plan is to have analytics as a central part of strategy…the early success stories and the overall trend has made the company re-consider the importance of analytics in strategy making.*"
>
> (*Avicom*)

While in some cases such adaptations to management structure and strategy orientation may be a result of successful outcomes, there are many reports in which benefits from big data analytics are not easy to quantify or demonstrate (Ransbotham, Kiron, & Prentice, 2016), leading to a lock-in in terms of the importance that big data analytics may have in driving organizational strategy.

4.3.2. Organizational inertia

A prominent theme from the case studies was the inertia encountered when attempting to implement data-driven decision-making in their companies. Resistance to change was encountered at multiple levels within their organizations, and at different phases of implementation. All three cases noted that while maturing their big data analytics capabilities is one side of the story, the other is the resistance faced when doing so and the tendency to fall back to previous ways of making decisions. Their responses also indicated that there are different levels at which inertia is present, which in many cases originates from the top management. Respondents stressed that this was a serious issue that largely influenced success of projects:

> "*We were met with skepticism five years ago when we proposed to management to experiment with big data analytics…this was before the trend had started and they could not see what the value would be…we had some reporting based on data warehousing and this was considered as sufficient…it took a lot of time and effort to try to convince management that we needed to invest in big data analytics*"
>
> (*Avicom*)

The respondents stated that top management overcame their concerns about investing in big data analytics mainly due to mimetic pressures since some of the forerunners in their industries had already done so. Other forms of inertia were present in inter-departmental collaborations when it came to big data analytics projects. In many cases organizations faced the problem of siloed data, which is a deeper issue of isolated departments. In the case studies this problem was particularly evident for the two large organizations where departments had their own line manager and had little inter-departmental coordination. Negative psychology and fear of losing their authority is a cause that many organizations, especially larger and more fragmented ones, erect barriers for cooperation with analytics departments. This creates large problems since in most cases analytics initiatives require data and domain knowledge from employees that are from different departments. The respondent from GLM noted this independent form of work in the following quote:

> "*We do not work together with other departments such as Risk and Credit. In essence, we do not cooperate besides for issues concerning common infrastructure* (*i.e. which platforms to use*). *We make decision and provide specifications and requirements to the IT department which takes all necessary actions to cover our needs*"
>
> (*GLM*)

Similar findings were noted at several points during the interviews and hint that achieving big data analytics capability maturity necessitates a solid understanding of how analytics are deployed in the organizational context. This will allow for the development of detailed deployment plans which foresee such obstacles and present solutions. The findings from the quantitative analysis pinpoint that such barriers may be overcome by establishing solid big data governance schemes which dictate how departments should cooperate and set roles and responsibilities.

4.3.3. Ethics and legislation

The importance of ethics and legislation was mentioned in all three case studies, as well as the implications they create for conducting analytics projects that lead to business value. All three organizations operate in Europe and are therefore subject to the General Data Protection Regulation (GDPR). While the GDPR was not mentioned as a barrier, it was considered as an opportunity since respondents believed that other organizations would have problems complying with the directives. Specifically, one interviewee from DataCom mentioned the following:

> "*GDPR can be considered as a barrier when it comes to the profiling of users through social media analytics. Since we do not apply profiling analysis to our data… the GDPR can be considered as a threat for some of our competitors and an opportunity for us.*"
>
> (*DataCom*)

Also, the ethics of using big data analytics and the repercussions it may have to the organization was a topic that was mentioned. Respondents noted that it is important for firms not only to comply with legislation, but also follow ethical rules when collecting, managing, and analyzing data. What was noted here is that while legislations such as the GDPR may provide guidelines about how to treat and manage data, they do not specify what constitutes an ethical decision about what type of insight you can extract and how you can use this knowledge. The respondents noted that it is important that firms build an image as a trustworthy entity for their customers to consent to provide data and allow them to leverage this data appropriately and within what they believe is an ethically correct approach. Specifically, the respondent from the GLM noted the following:

> "*People become increasingly aware in how their personal data are being used. The recent Facebook personal data* "*abuse*" *story has contributed to this phenomenon. On the one hand we all want to make profit by analyzing personal data. On the other hand, if you think, for example, that after our meeting I will get a Linkedin message to connect with you, it is somewhat scaring. So, yes there is an ethical issue in the use of private data that may raise a barrier in our business the near future. It all depends in the value we give back to our customers by analyzing their data. For example, I totally agree to give personal data in order to get relevant content. People must see a value in it.*"
>
> (*GLM*)

5. Discussion

While the hype around big data analytics is continuously growing, the conditions under which such investments lead to business value remain largely unexplored in empirical research. The value of big data analytics has also been questioned in recent articles, since it is argued that only a small percentage of companies have been able of capture the full potential of their big data investments (Ross, Beath, & Quaadgras, 2013). This fact is rather striking when considering the vast number of articles of business publications that talk about the transformative power of big data analytics. Gupta and George (2016) argue that this phenomenon can be largely attributed to the fact that most of the literature on big data analytics has been drafted by consultants, therefore lacking in theoretical grounding and large-scale empirical testing. They also note that what is important is not the technologies surrounding big data analytics, but rather, the organizational diffusion of such technologies towards the generation of insight that can help attain business goals. Most studies to date argue that organizations should invest in certain key areas and establish specific processes and practices to realize value from their big data investments. The objective of this study is to understand if big data analytics can lead to any measurable business value and to isolate the core resources in different context. We question the assumption that all organizations require the same approach when deploying big data analytics to support their business strategies. To address this research question, we isolate the resources that comprise a big data analytics capability and are argued to be necessary for firms to realize business value. Our research is grounded on a mixed-methods approach bridging quantitative and qualitative methods. By applying a fsQCA research method on a sample of 175 Greek firms, we empirically demonstrate that there are four main clusters of firms that represent different combinations of core elements in their attainment of performance gains from big data analytics. These differences are argued to be a result of the different contexts in which these firms operate, showing that there is equifinality in achieving performance gains. To understand how these factors coalesce, as well as to uncover additional elements that may enable or impede business value, we conducted three case studies with companies that had experience in deploying big data analytics to support their operations.

5.1. Implications for research

From a theoretical perspective, the findings of this study add to existing literature in several ways. First, they demonstrate how a complexity approach can be empirically explored in the context of big data analytics. Such approaches are seldom investigated in quantitative studies with regards to big data analytics and business value. The main tenets of complexity theory have been useful in explaining phenomena in a number of different of domain areas, including those of information systems management, and provide interesting insight about how certain outcomes emerge and particularly the combinatory and non-linear effect that antecedents may have on them (El Sawy et al., 2010; Mikalef & Pateli, 2017; Woodside, 2014). One of the main issues faced with many organizations that adopt big data analytics is that they are failing to realize performance gains. Literature has attributed this to the fact that not all organizations operate under the same conditions and therefore require a different approach concerning how they invest and deploy their big data analytics resources (Abbasi et al., 2016). Complexity theory and the corresponding reach method enable us to understand how contextual factors shape the importance of certain resources and jointly lead to performance gains. The move towards such theoretical and methodological approaches is advocated to lead to the generation of novel theoretical and practical implications (Woodside, 2013a).

Second, despite much anecdotal claims concerning the enabling effect that big data analytics have on strengthening existing or realizing performance gains, there is still limited empirical research to consolidate them. Our findings show that different combinations of big data-related resources have a greater or lesser significance depending on the context they are used in. More precisely, we find that more technological and technical resources contribute towards performance gains in moderately uncertain environments, while organizational aspects and managerial skills are of greater importance in highly uncertain conditions. In fact, there are several business reports that argue about the importance that organizational aspects may have in realizing business value (Kiron, 2017), but very limited empirical evidence to confirm such claims and demonstrate what combinations of factors lead to performance gains. Our findings show that big data analytics should not be perceived as a solely technical challenge, but rather, an organizational one which requires fusion with the firm's business strategy. Therefore, understanding the constituent components that enable such a fusion between big data analytics and business strategy, and that as a result lead to performance gains, is critically important.

The findings of this study differ significantly from existing literature in the area of big data analytics and business value and show that different factors must be emphasized depending on the context of examination. To date, most studies build on the assumption that firms need to follow the same approach when investing in big data analytics, and therefore in their majority, do not distinguish between different types of organizations and the contexts in which they operate. Furthermore, through the three case studies we see that there are several aspects that should be considered when deploying these resources, and that achieving maturity in each requires an additional set of elements to be taken into account. We uncover three main themes that were critical according to the organizations we used as case studies, which result in several research directions that can serve as the basis for future research. Our findings show that the ways in which strategies surrounding big data analytics are developed and executed depend on several factors, including the size-class, the organizational structure, the industry as well as on top manager support. Furthermore, they point out to several other aspects that are likely to emerge as key competitive differentiators in the near future, such as the ethical issues surrounding use of big data and the trust developed between firms and their customers.

5.2. Implications for practice

The results of this study also have several interesting implications for practitioners. First, this study shows that big data analytics is more than just mere investments in technology, collection of vast amounts of data, and allowing the IT department to experiment with novel analytics techniques. Complementary to the above-mentioned, important elements of gaining business value out of big data investments include recruiting people with good technical and managerial understanding of big data and analytics, fostering a culture of organizational learning, and embedding big data decision-making into the fabric of the organization. Hence, it is the combined effect of these resources that will enable a firm to develop a big data analytics capability and realize value gains. This of course means that a multitude of processes need to be put into action, which requires top management commitment and a clear plan for firm-wide big data analytics adoption and diffusion. A number of studies have already begun to highlight the significance of all these factors, and provided managers with guidelines on how to develop and mature their big data analytics capabilities (Hindle & Vidgen, 2018; Mikalef, Framnes, et al., 2017; Vidgen et al., 2017).

By outlining the core resources that are needed to develop a big data analytics capability, this study can help managers construct an assessment tool, so they can benchmark their organizations strengths and weaknesses. The main pillars can help expose areas that have been under-developed or insufficiently funded. Resources of an intangible nature, such as a data-driven culture and governance practices, can provide managers with an understanding of the importance of these aspects, and help them form strategies to strengthen them throughout the firm. Given that many companies are still at an inaugurating stage in their big data analytics initiatives, it is critical to have a good overview of all the areas that should be considered in connection to realize value, as well as to calculate expected costs and gains. Furthermore, while some resources such as technical, data, and even human skills can be quite easily and quickly replicated or acquired from the market, others, such as a data-driven culture require planning and a well-documented process to form and mature. Hence, an additional practical implication concerns the calculation of the time and complexity that some resources require to develop.

Managers should therefore think about the maturation time required well before they expect any measurable outcomes from their big data investments. Adding to this, recruiting employees with the technical and managerial skills necessary in the age of big data is a great concern for many executives. Our findings showcase the importance of these in realizing business value. Furthermore, our case studies demonstrate some issue managers often tend to overlook, that of inertia during diffusion of technological innovations. The main premise that big data analytics build on is that the insight generated by big data analytics will be used to transform operations and lead to enhanced ways of capturing value. Nevertheless, inhibiting forces were reported to obstruct both the successful diffusion of big data analytics within firm boundaries, but also in the capturing of opportunities after insight had been generated. This poses a critical issue for managers who must develop appropriate mechanisms and practices to overcome such barriers.

5.3. Limitations and future research

Despite the contributions of the present study it is constrained by a number of limitations that future research could seek to address. First, as noted already, self-reported data are used to test our research questions. Although considerable efforts were undertaken to confirm data quality, the potential of biases cannot be excluded. The perceptual nature of the data, in combination with a study-design that uses a single key informant, could suggest that there is bias, and that factual data do not coincide with respondents' perceptions. Despite this, relying on top management respondents as key informants is a good way to minimize bias, as they typically have good knowledge on various related domains. Future studies could follow an alternative approach by sampling multiple respondents within a single firm since that would be a useful way to establish inter-rater validity and to improve internal validity. Second, although we examine the effect of resources related to big data analytics on firm performance, we do not factor in several other important contextual factors. It is highly probable that the value of directing big data initiatives may be more beneficial in some cases than in others or be dependent on the time-frame since they have been deployed. This is an area that future research should seek to address, and it is of increased practical value, particularly considering the costs of deploying big data analytics. It is important to understand in each industry how big data analytics capabilities are developed, as well as through what mechanisms they produce value, and how that can be captured. Third, one of the limitations of the research method used is that we cannot include the notion of time in our analysis, meaning the length that it takes organizations to acquire, orchestrate, and deploy these big data analytics resources to achieve performance gains. Future studies looking at multiple case studies and following a longitudinal approach could seek to uncover the process of deploying such resources and the barriers that are faced during different phases.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 704110.

Appendix A. Survey instrument

Measure	Item
Data	
	D1. We have access to very large, unstructured, or fast-moving data for analysis
	D2. We integrate data from multiple sources into a data warehouse for easy access
	D3. We integrate external data with internal to facilitate analysis of business environment
Technology	
	T1. We have explored or adopted parallel computing approaches (e.g., Hadoop) to big data processing
	T2. We have explored or adopted different data visualization tools
	T3. We have explored or adopted new forms of databases such as Not Only SQL(NoSQL)
	T4. We have explored or adopted cloud-based services for processing data and performing analytics
	T5. We have explored or adopted open-source software for big data analytics
People	
Managerial skills	MS1. Our BDA managers are able to understand the business need of other functional managers, suppliers, and customers to determine opportunities that big data might bring to our business.
	MS2. Our DBA managers are able to coordinate big data-related activities in ways that support other functional managers, suppliers, and customers
	MS3. Our BDA' managers are able to understand and evaluate the output extracted from big data
	MS4. Our BDA' managers are able to understand where to apple big data
Technical skills	TS1. Our 'big data analytics' staff has the right skills to accomplish their jobs successfully
	TS2. Our 'big data analytics' staff is well trained
	TS3. We provide big data analytics training to our own employees
	TS4. Our 'big data analytics' staff has suitable education to fulfil their jobs
Organization	
Structural practices	In our organization, we ____________
	STR1. have identified key IT and non-IT decision makers to have the responsibility regarding data ownership, value analysis and cost management.
	STR2. use steering committees to oversee and assess data values and costs
Relational practices	In our organization, we ____________
	RLT1. educate users and non-IT managers regarding storage utilization and costs
	RLT2. develop communications regarding policy effectiveness and user needs
Data-driven culture	DD1. We base our decisions on data rather than on instinct
	DD2. We are willing to override our own intuition when data contradict our viewpoints
	DD3. We continuously coach our employees to make decisions based on data
Process	
Procedural practices	In our organization, we have controlled practices regarding data management in terms of ____________
	PCR1. setting retention policies (e.g. time to live) of data
	PCR2. backup routines
	PCR3. establishing/monitoring access (e.g. user access) to data
	PCR4. classifying data according to value
	PCR5. monitoring costs versus value of data
Performance	We perform much better than our main competitors in terms of:
	PER1. Profitability
	PER2. Profits as percentage of sales
	PER3. Decreasing product or service delivery cycle time
	PER4. In reducing operating costs
	PER5. In profit growth rates
	PER6. Rapid response to market demand
	PER7. Rapid confirmation of customer orders
	PER8. Increasing customer satisfaction
	PER9. Providing better product and service quality
	PER10. In reducing operating costs
Dynamism	With respect to the uncertainty of your environment, please indicate how much you agree or disagree with the following statements
	DYN1. Products and services in our industry become obsolete very quickly
	DYN2. The product/services technologies in our industry change very quickly
	DYN3. We can predict what our competitors are going to do next (Reverse coded)
	DYN4. We can predict when our products/services demand changes (Reverse coded)
Heterogeneity	With respect to the uncertainty of your environment, please indicate how much you agree or disagree with the following statements
	HET1. Customer buying habits
	HET2. Nature of competition
	HET3. Product lines
Hostility	With respect to the uncertainty of your environment, please indicate how much you agree or disagree with the following statements
	HOS1. Scarce supply of labor
	HOS2. Scarce supply of materials
	HOS3. Tough price competition
	HOS4. Tough competition in product/service quality
	HOS5. Tough competition in product/service differentiation

Appendix B. Heterotrait-monotrait ratio (HMTM)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Managerial skills										
(2) Technical skills	0.436									
(3) Data-driven culture	0.371	0.472								
(4) Procedural practices	0.320	0.419	0.326							
(5) Relational practices	0.381	0.285	0.342	0.382						
(6) Structural practices	0.402	0.401	0.305	0.421	0.432					
(7) Performance	0.421	0.402	0.351	0.358	0.435	0.503				
(8) Dynamism	0.245	0.387	0.420	0.378	0.343	0.204	0.416			
(9) Heterogeneity	0.275	0.661	0.470	0.333	0.376	0.296	0.286	0.225		
(10) Hostility	0.329	0.358	0.482	0.267	0.312	0.257	0.424	0.350	0.377	

Appendix C. Fuzzy set calibration

Variable	Mean (S.D.)	Percentiles			Thresholds		
		25%	50%	75%	Full membership	Cross-over point	Full non-membership
Data	4.98 (1.72)	3.22	4.95	6.08	6.08	4.95	3.22
Technology	4.61 (2.02)	3.17	4.51	5.84	5.84	4.51	3.17
Technical skills	4.51 (1.82)	3.31	4.50	5.59	5.59	4.50	3.31
Managerial skills	5.07 (1.84)	3.45	5.02	5.97	5.97	5.02	3.45
Structural practices	4.45 (1.95)	3.06	4.37	5.62	5.62	4.37	3.06
Relational practices	4.10 (1.51)	3.02	4.08	5.24	5.24	4.08	3.02
Data-driven culture	5.01 (1.81)	3.42	4.98	6.03	6.03	4.98	3.42
Procedural practices	5.03 (1.82)	3.48	4.97	5.95	5.95	4.97	3.48
Performance	3.94 (1.39)	2.89	3.90	5.32	5.32	3.90	2.89
Dynamism	4.67 (1.45)	3.47	4.45	5.53	5.53	4.45	3.47
Heterogeneity	4.13 (1.34)	3.12	4.07	5.11	5.11	4.07	3.12
Hostility	4.79 (1.64)	3.51	4.75	5.32	5.32	4.75	3.51

Appendix D. Interview guidelines

1) Introduction

1a Would you tell us about what your organization does?
1b Would you tell us about what you do and your background?

2) Big data context

2a What types of data are held by your organization? Of this data, what would you consider to be "big data"?
2b Who uses the data?
2c Who owns the data (manages add/change/delete)?
2d What big data technologies are used in your organization?
2e. Who first introduced these tools to the organization (from within IT, from another function)?
2f. Who uses these tools?
2g. Are these tools managed locally or in the cloud (e.g. Amazon AWS).

3) Big data value creation

3a How do big data workers create value in the company? How do they influence decision-making? What types of decisions?
3b Is value created internally (e.g., to improve customer retention) or externally (e.g., to sell data products)?
3c How is the business value of your big data evaluated (if at all)?

4) Organization

4a How many people with deep data skills work in your company?
4b Do you have any formal strategy in the company in using big data analytics?
4c Is there is any governance scheme to operationalize this strategy? If yes, what does it entail?
4d How are these people referred to in the company? What is their job role?
4e In what part of the company do they work? Are they part of a cross-functional team or are they generally found operating within a single corporate function?
4f Are they generalists or domain specialists?
4g What types of activities are they involved with?

4h What proportion of data analysis is exploratory (e.g., Google day)? How is this justified in business terms?
4i What analytics techniques do they use? E.g., machine learning, recommendation systems, sentiment analysis, time series analysis, SNA, AI, simulations.
4j What desktop analytics and visualization technologies do they use? E.g., Excel, SAS, Stata, SPSS, R, Python, Mahout, Tableau?
4k Does the output from these tools feed management reporting? If so, who is the end consumer of the reporting from these outputs?

5) Process

5a Could you describe the lifecycle of a 'big data project' (e.g. a recent project)?
5b What are the best ways to motivate data workers in your company? How do they differ, in this respect, from other workers?
5c Who sets specific roles and tasks related to big data analytics? Can you give some examples of policies implemented?
5d How self-directed is their work? Do they generate new questions or answer questions posed by others?
5e What is the single thing your company has done that has made your data workers more productive?
5f What is the main barrier to generating more value from them?
5g Do you believe that there is a data-driven culture in your company? What factors have contributed in enabling/hindering this culture? What could be improved?

6) People

6a What knowledge/skills/competences are you looking for when you hire them?
6b Where do you go to look for them (from universities/from industry/elsewhere)?
6c How long does it take it for them to be able to make a contribution to the business?
6d What training and development is given to big data workers?
6e What are the career progression opportunities for big data workers?
6f If they leave the organization, where do they go and why?
6g Are there any particular skill sets or competences in short supply in the market? Are these getting worse or better?
6h What is the impact of these shortages (if they exist)?
6i What are you doing to address these shortages?
6j What skills and competencies should Universities be developing in its analytics graduates?
6k What future skills do you project to be important?

7) Barriers

7a What regulatory or legal constraints are there? Do these inhibit what can be done with big data?
7b What ethical issues are you concerned about (these might be legal but not something you would want to do).
7c Going forward, what other barriers to using and creating value from big data do you see? E.g., organizational, managerial, technical.

8) Value

8a What would you say is the value of big data analytics for your company?
8b Have you been able to capture this value? Is it easy to quantify?
8c What are some of the main obstacles you encountered in the implementation of the analytics project?
8d In which areas has big data analytics benefited you most? (e.g. products, services, processes? Or in which way, radical or incremental improvements).
8e How has the environment in which you compete influenced your decision to adopt data analytics?

References

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems, 17*(2).

Anderson, P. (1999). Perspective: Complexity theory and organization science. *Organization Science, 10*(3), 216–232.

Bechor, T., Neumann, S., Zviran, M., & Glezer, C. (2010). A contingency model for estimating success of strategic information systems planning. *Information & Management, 47*(1), 17–29.

Boudreau, M.-C., Gefen, D., & Straub, D. W. (2001). Validation in information systems research: A state-of-the-art assessment. *MIS Quarterly,* 1–16.

Cabrera, D., Cabrera, L., Powers, E., Solin, J., & Kushner, J. (2018). Applying systems thinking models of organizational design and change in community operational research. *European Journal of Operational Research, 268*(3), 932–945.

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

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4).

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology, 30*(1), 44–57.

Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods, 4*(2), 144–-192.

El Sawy, O. A., Malhotra, A., Park, Y., & Pavlou, P. A. (2010). Research commentary—Seeking the configurations of digital ecodynamics: It takes three to tango. *Information Systems Research, 21*(4), 835–848.

Farrell, A. M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research, 63*(3), 324–327.

Fiss, P. C. (2007). A set-theoretic approach to organizational configurations. *Academy of Management Review, 32*(4), 1180–1198.

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal, 54*(2), 393–420.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems, 26*(3), 191–209.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management, 53*(8), 1049–1064.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice, 19*(2), 139–152.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135.

Hindle, G. A., & Vidgen, R. (2018). Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research, 268*(3), 836–851.

Kiron, D. (2017). Lessons from becoming a data-driven organization. *MIT Sloan*

Management Review, 58(2).
Kooper, M. N., Maes, R., & Lindgreen, E. R. (2011). On the governance of information: Introducing a new concept of governance to support the management of information. *International Journal of Information Management, 31*(3), 195–200.
Kumar, N., Stern, L. W., & Anderson, J. C. (1993). Conducting interorganizational research using key informants. *Academy of Management Journal, 36*(6), 1633–1651.
LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review, 52*(2), 21.
Leavitt, H. J. (1965). Applied organizational change in industry, structural, technological and humanistic approaches. *Handbook of organizations* (pp. 264). .
Liu, H., Ke, W., Wei, K. K., & Hua, Z. (2013). The impact of IT capabilities on firm performance: The mediating roles of absorptive capacity and supply chain agility. *Decision Support Systems, 54*(3), 1452–1462. https://doi.org/10.1016/j.dss.2012.12.016.
MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly,* (2), 293–334.
Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*.
Mas-Verdú, F., Ribeiro-Soriano, D., & Roig-Tierno, N. (2015). Firm survival: The role of incubators and business characteristics. *Journal of Business Research, 68*(4), 793–796.
McAfee, A., Brynjolfsson, E., & Davenport, T. H. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 60–68.
Mendel, J. M., & Korjani, M. M. (2012). Charles Ragin's fuzzy set qualitative comparative analysis (fsQCA) used for linguistic summarizations. *Information Sciences, 202*, 1–23.
Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. *British Journal of Management*. https://doi.org/10.1111/1467-8551.12343 (In press).
Mikalef, P., Framnes, V. A., Danielsen, F., Krogstie, J., & Olsen, D. H. (2017). *Big data analytics capability: antecedents and business value. Paper presented at the Pacific Asia Conference on Information Systems, Langkawi, Malaysia.* 2018.
Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2017). Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and e-Business Management,* 1–32.
Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and e-Business Management, 16*, 1–32.
Mikalef, P., & Pateli, A. (2017). Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA. *Journal of Business Research, 70*, 1–16.
Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization, 17*(1), 2–26.
Navarro, S., Llinares, C., & Garzon, D. (2016). Exploring the relationship between co-creation and satisfaction using QCA. *Journal of Business Research, 69*(4), 1336–1339.
Newkirk, H. E., & Lederer, A. L. (2006). The effectiveness of strategic information systems planning under environmental uncertainty. *Information & Management, 43*(4), 481–501.
Nunnally, J. (1978). *Psychometric methods.* New York: McGraw-Hill.
Ordanini, A., Parasuraman, A., & Rubera, G. (2014). When the recipe is more important than the ingredients: A qualitative comparative analysis (QCA) of service innovation configurations. *Journal of Service Research, 17*(2), 134–149.
Pappas, I. O., Mikalef, P., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2018). Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. *16*(3), 479–491.
Park, Y., El Sawy, O. A., & Fiss, P. C. (2017). The role of business intelligence and communication technologies in organizational agility: A configurational approach. *Journal of the Association for Information Systems, 18*(9).
Petter, S., DeLone, W., & McLean, E. R. (2013). Information systems success: The quest for the independent variables. *Journal of Management Information Systems, 29*(4), 7–62.
Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly, 31*(4), 623–656.
Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879.
Popovič, A., Hackney, R., Tassabehji, R., & Castelli, M. (2018). The impact of big data analytics on firms' high value business performance. *Information Systems Frontiers, 20*(2), 209–222.
Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health information science and systems, 2*(1), 3.
Ragin, C. C. (2009). Qualitative comparative analysis using fuzzy sets (fsQCA). *Configurational comparative methods, 51*.
Ragin, C. C., & Fiss, P. C. (2008). Net effects analysis versus configurational analysis: An empirical demonstration. *Redesigning social inquiry: Fuzzy sets and beyond,* 190–212.
Rai, A., & Tang, X. (2010). Leveraging IT capabilities and competitive process capabilities for the management of interorganizational relationship portfolios. *Information Systems Research, 21*(3), 516–542.
Ransbotham, S., & Kiron, D. (2017). Analytics as a source of business innovation. *MIT Sloan Management Review, 58*(3), 1–16.
Ransbotham, S., Kiron, D., & Prentice, P. K. (2016). Beyond the hype: The hard work behind analytics success. *MIT Sloan Management Review, 57*(3).
Ringle, C. M., Wende, S., & Becker, J.-M. (2015). SmartPLS 3. Boenningstedt: SmartPLS GmbH. http://www.smartpls.com.
Ross, J. W., Beath, C. M., & Quaadgras, A. (2013). You may not need big data after all. *Harvard Business Review, 91*(12), 90.
Schneider, C. Q., & Wagemann, C. (2010). Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets. *Comparative Sociology, 9*(3), 397–418.
Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems, 23*(4), 433–441.
Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *The Communications of the Association for Information Systems, 13*(1), 63.
Tallon, P. P. (2013). Corporate governance of big data: Perspectives on value, risk, and cost. *Computer, 46*(6), 32–38.
Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The information artifact in IT governance: Toward a theory of information governance. *Journal of Management Information Systems, 30*(3), 141–178.
Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly, 37*(1).
Verdu, A. J., Tamayo, I., & Ruiz-Moreno, A. (2012). The moderating effect of environmental uncertainty on the relationship between real options and technological innovation in high-tech firms. *Technovation, 32*(9–10), 579–590.
Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research, 261*(2), 626–639.
Vom Brocke, J., Schmiedel, T., Recker, J., Trkman, P., Mertens, W., & Viaene, S. (2014). Ten principles of good business process management. *Business Process Management Journal, 20*(4), 530–548.
Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics, 34*(2), 77–84.
Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356–365.
Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics, 176*, 98–110.
White, L. (2018). A Cook's tour: Towards a framework for measuring the social impact of social purpose organisations. *European Journal of Operational Research, 268*(3), 784–797.
Wilden, R., Devinney, T. M., & Dowling, G. R. (2016). The architecture of dynamic capability research identifying the building blocks of a configurational approach. *The Academy of Management Annals, 10*(1), 997–1076.
Woodside, A. G. (2012). Proposing a new logic for data analysis in marketing and consumer behavior: Case study research of large-N survey data for estimating algorithms that accurately profile X (extremely high-use) consumers. *Journal of Global Scholars of Marketing Science, 22*(4), 277–289.
Woodside, A. G. (2013a). Moving beyond multiple regression analysis to algorithms: Calling for adoption of a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory. *Journal of Business Research, 66*(4), 463–472.
Woodside, A. G. (2013b). *Moving beyond multiple regression analysis to algorithms: Calling for adoption of a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory.* Elsevier.
Woodside, A. G. (2014). Embrace•perform•model: Complexity theory, contrarian case analysis, and multiple realities. *Journal of Business Research, 67*(12), 2495–2503.
Wu, P.-L., Yeh, S.-S., & Woodside, A. G. (2014). Applying complexity theory to deepen service dominant logic: Configural analysis of customer experience-and-outcome assessments of professional services for personal transformations. *Journal of Business Research, 67*(8), 1647–1670.
Yin, R. K. (2009). *Case study research: Design and methods (applied social research methods).* London and Singapore: Sage.
Yin, R. K. (2017). *Case study research and applications: Design and methods.* Sage publications.

Patrick Mikalef is a Marie Skłodowska-Curie post-doctoral research fellow in the area of Information Systems Strategy. He received his B.Sc. in Informatics from the Ionian University, his M.Sc. in Business Informatics for Utrecht University, and his Ph.D. in IT Strategy from the Ionian University. His research interests are on strategic use of information systems and IT-business value in turbulent environments. He has published work in international conferences and peer-reviewed journal including the Journal of Business Research, Industrial Management & Data Systems, Journal of Theoretical and Applied Electronic Commerce Research, and Health Information and Libraries Journal.

Maria Boura is post-doctoral researcher at the Department of Management Science and Technology of the Athens University of Economics and Business. Her research interests focus on Business Strategy, Corporate Social Responsibility, Business Ethics and Corruption, and Big data and management. Her research papers have been published in international academic conferences (Academy of Management, European Academy of Management, British Academy of Management, European Group for Organizational Studies, European Conference on Information Systems). She holds a PhD and an MSc from the Athens University of Economics and Business (Greece).

George Lekakos is Associate Professor at the Department of Management Science and Technology, Athens University of Economics and Business, Greece. He is also Director of the MSc in Management Science and Technology. He leads the Intelligent Media Lab (IML), which is a research group within the ELTRUN Research Center (http://www.eltrun.gr) at the Athens University of Economics and Business (AUEB), Greece. His research focuses on Machine learning, Recommender Systems, Big Data and Strategy, and e-business. He has published more than sixty papers in international journals and conferences (e.g. EJOR, International Journal of Electronic Commerce, User Modelling and

User adapted interaction etc.), and he is co-editor of books, conference proceedings, journals' special issues and serves as editorial board member of international journals.

John Krogstie holds a PhD (1995) and a MSc (1991) in information systems from the Norwegian University of Science and Technology (NTNU), where he is currently a full professor in information systems at the IDI-department. At IDI he is Department Head. John Krogstie is the Norwegian representative and Vice-Chair for IFIP TC8 and was chair of IFIP WG 8.1 on information system design and evaluations (2010–2015). His research interests are information systems modelling, quality of models and modelling languages, eGovernment and mobile information systems. He has published around 250 refereed papers in journals, books and archival proceedings since 1991. H-index as of July 2018 is 40, G-index 58 according to Google Scholar.