



# C

## ONCEPTUALIZATION AND THEORIZATION OF THE BIG DATA

<sup>1</sup>Marcos Mazieri  
<sup>2</sup>Eduardo Dantas Soares

### ABSTRACT

The term Big Data is being used widely by companies and researchers who consider your relevant functionalities or applications to create value and business innovation. However, some questions arise about what is this phenomenon and, more precisely, how it occurs and under what conditions it can create value and innovation in business. In our view, the lack of depth related to the principles involved in Big Data and the very absence of a conceptual definition, made it difficult to answer these questions that have been the basis for our research. To answer these questions we did a bibliometric study and extensive literature review. The bibliometric studies were realized based in articles and citation of Web of Knowledge database. The main result of our research is the providing a conceptual definition for the term Big Data. Also, we propose which principles discovered can contribute with other researches that intend value creation by Big Data. Finally we propose see the value creation through Big Data using the Resource Based View as the main theory used for discuss that theme.

**Keywords:** Big Data; Innovation; Business Model; Business Innovation; Review Study; Resource Based View (RBV).

---

<sup>1</sup> Digitalplace Network Corp. United States. [m\_mazzieri@hotmail.com]

<sup>2</sup> Doctoral candidate at Nove de Julho University (UNINOVE), São Paulo – SP (Brazil). Currently, he is working at Faculdade Nossa Cidade, São Paulo (Brazil). [edudantassp@gmail.com]

# C ONCEITO E TEORIA DE BIG DATA

## RESUMO

O termo Big Data está sendo amplamente utilizado por empresas e pesquisadores que consideram suas funcionalidades ou aplicações relevantes para criar valor e inovação empresarial. No entanto, algumas perguntas surgem sobre o que é este fenômeno e, mais precisamente, como ele ocorre e em que condições ela pode criar valor e inovação nos negócios. Em nossa opinião, a falta de profundidade relacionada com os princípios envolvidos do Big Data e a própria ausência de uma definição conceitual, tornou difícil responder a estas questões que serviram de base para nossa pesquisa. Para responder a estas perguntas fizemos um estudo bibliométrico e uma extensa revisão da literatura. Os estudos bibliométricos foram realizadas com base em artigos e citações da base *Web of Knowledge*. O principal resultado da nossa pesquisa é o fornecimento de uma definição conceitual para o termo Big Data. Além disso, propomos que os princípios descobertos podem contribuir com outras pesquisas que pretendem investigar a criação de valor por Big Data. Finalmente propomos que a criação de valor por meio do uso do Big Data deve ser vista à luz da Visão Baseada em Recursos, pois essa é a principal teoria usada para discutir esse tema.

**Palavras-chave:** Big Data, Inovação, Geração de Valor, Modelo de Negócios, Inovação Empresarial, Visão Baseada em Recursos.

## INTRODUCTION

The significant increase in capacity creation and storage of data, made possible by the start of the digital era in 2000, provided the first speculation about a phenomenon known as Big Data (Hilbert & Lopez, 2011). The descriptions of the elements that make up the term Big Data are common in the literature, and are usually explained by assets formed by information. Characterized by having a large volume of data that form where the speed with which they are formed and the variety that acquire, require specific technologies, and own analytical methods for processing value (Andrea De Mauro, 2014).

The term Big Data has emerged initially in 2001 as assets formed by high-volume information, speed and variety (Laney, 2001). After that, there were some other definitions of the term (see Baaziz & Quoniam, 2013; Beyer, 2011; Hilbert & Lopez, 2011; Silveira, Marcolin, & Freitas, 2016; Taurion, 2013), as well as criticism and doubts about the relevance of the subject as something to be dealt with by science or just a fad (Buhl, Röglinger, Moser, & Heidemann, 2013). Even with these propositions, we find that there seems to be a lack of a more robust

definition of Big Data, which can provide a better understanding of the concept. This lack of definition may lead to qualify point any large database and Big Data. Such inaccuracies could pose problems in the recognition of the value of data and constantly hinder the creation of value and innovation for business. This is the argument that we support and that can be read in the "discussion" in this chapter.

When we seek the repositories of scientific and technical publications - Google Scholar, Web of Science and Espacenet - studies describing the phenomenon Big Data, we found that these works are constructed favoring the description of the elements that make up the environments for treatment of Big Data and predominantly published by consulting and information technology manufacturers for Big Data.

Normally serves to structural description of the methods that involve the processing of data, the description of their potential, their architectural and functional description, and software and hardware

elements involved in the implementation of imaging systems, storage and analysis.

Our current study theoretical, descriptive, inductive, predominantly quantitative, supports a new proposal for the definition of Big Data.

We are motivated to do this work because we checked into our bibliometric study on the subject, which is not clear conceptual definition of Big Data with phenomenon and argue that this obscurity may be hindering or preventing the appropriation of knowledge on the subject, due to ignorance of the principles involved in this phenomenon. The lack of conceptual clarity of the principles involved in Big Data, prevent are understood the possible consequences of the use of Big Data, including the creation of value and innovation.

Additionally, based on bibliometric study and content analysis of technical reports and papers, we intend to propose some theoretical discussions regarding the creation of value and innovation for business using mining in Big Data. Nevertheless, by treat about a subject where there is much controversy there against points that have doubts about the euphoria involved in matters related to Big Data.

The main criticisms relate treat yourself to just another fad, or an ongoing advertising campaign to stimulate the market of manufacturers and information technology suppliers as analyzing data. It is something done since the 1960s, not showing nothing new about that (Buhl, 2013).

Discussions will also step into the right areas, of which we highlight those that represent obstacles to the advancement of the use of Big Data; national laws that protect the privacy of consumers and people, has not been discussed with the necessary depth. We agree that this is something new and that the use of Big Data to create value and innovation for business, can not be achieved automatically.

Buhl at al. (2013), confirms that to exploit Big Data, should be considered the volume of challenges, speed, range, veracity and privacy. This last challenge, privacy, was presented as the element that can derail the realization of the expected potential for Big Data, especially difficult integration between big data and business models. Although we agree with the placement of Buhl (2013), our paper argues that the findings of Buhl et al. (2013) only considered the level of operational analysis, that it is important, it seems to require a reinterpretation, considering the conceptual definition - not operationally, as we are trying to make this our article.

Basing on the scientific literature identified in the bibliometric study presented in the next section, it was built our literature review, followed by a description of the methodological procedures.

Sequentially, we present the results achieved, especially focused on content analysis of articles evaluated and proposed experiments, then passing to the discussion

of the results of both the content analysis, the experiment, which justify the conceptual definition of Big Data and provided views on possible future studies. Close work with the "conclusion."

## **BIBLIOMETRIC STUDIES UPON INNOVATION THROUGH BIG DATA**

The fact that we are looking for a conceptual definition of Big Data, and operational definitions or tooling, motivated us to carry out a prior bibliometrics literature review. As taught in the Buhl et al. (2013), which we have seen in the literature is the management application more targeted. This explains the title of the work Buhl et al. (2013) and encourages researchers to wonder if Big Data is just a fad, or something that exists for a long time, but has now been renamed, with a new name and format to sell hardware and software. We are arguing that this is not a fad, that is, in our study; we showed that Big Data is not something that soon will no longer be seen in the setting of global innovation. On the contrary, our studies have led us to identify that there are possibilities for the passage of contributions from Big Data to the business models, creating value and innovation.

Reserve this section to give an indication of how we bibliometrics, which the software used to automate it, some details of the configuration of such programs and the main findings, which explain why some schools of thought understand the Big Data as a fad or market action. Below we will explain specifically bibliometric methodological procedures and presentation of findings with comments that we deem relevant for better understanding.

### **Data base of documents upon Big Data**

There are currently at least 20 major bases of scientific data. We chose to study this examination of the Thomson Reuters SCOPUS - Web of Science - for the realization of bibliometrics and Google Scholar to identify the technical and scientific publications available on the web surface.

Considering the most cited work, we find overlap 100% between the two databases, therefore, chose the web of science to make bibliometrics, for the convenience of the search features offered on the platform. This database is a repository of scientific, journalistic and technical publications that allows that tries for keywords that can be interspersed with Boolean operators and wildcard. In this article, we chose to use the search to texts written in English and from the search expression ("big \* and \* Date") and ("innov \*"). With the syntax, increased the results for all endings of words after the "big" prefix, and also obtained as a result, all the texts that have in their title or abstract, the words with different endings after the prefix "date" since they appear together, because we use

the Boolean function "and". As interest of the study is to better understand the Big Data in the context of innovation, we use another Boolean operator "and" to search only the texts where Big Data and innovation appear together, also with all the endings of the word innovation. As a result, we obtained 2300 articles, which applied two filters available in the Web of Science environment. We leave only scientific articles and are registered in the economy and business area. After applying the filter, they considered 298 items for realization of bibliometric study.

The purpose of bibliometric to analyze the quotes used in these 298 articles, to map the knowledge of the field. Therefore, we use the BibExcel software to generate the list of citations present in the articles found in the keyword search. Once you have identified the quote, we decided to

discard those that appeared only 2 times or less. We therefore consider the low frequency of quotes represented low relevance for the study, while maintaining the convenience sample of 34 authors. Based on these authors have developed a symmetrical array of quotes, to submit it to exploratory factor analysis - AFE -. During adjustment procedures of AFE, aiming initially adjust the KMO and esfericidade test Bartlett for: higher than 0.6, see Hair 2008 (Joseph, Bush, & Ortinau, 2008), were systematically removed from the model the authors who presented communalities smaller than 0.5. The authors removed were Nonaka 1995 Vargo 2004 Russon P 2011 Rust R 1995 Leisure D 2009 Moffitti K 2013, Choi H 2012, Butler D 2013 Hastie T 2009 Lohr S 2012, Hevner The 2004 Waller 2013 Ginsberg J, 2009, 2014 Leisure D KMO After adjustments, authors remaining 17 are shown in **table 1**.

Table 1- Communalities results, after fit of KMO in 0.676

<i>Communalities</i>		Initial	Extraction
<i>Barney J, 1991</i>		1.000	.603
<i>Bollen J, 2011</i>		1.000	.779
<i>Boyd D, 2012</i>		1.000	.852
<i>Chen H, 2012</i>		1.000	.746
<i>Cukier K, 2013</i>		1.000	.800
<i>Dean J, 2008</i>		1.000	.841
<i>George G, 2014</i>		1.000	.816
<i>Grant R, 1996</i>		1.000	.754
<i>Harris J, 2007</i>		1.000	.834
<i>Kogut B, 1992</i>		1.000	.630
<i>Lavalle S, 2011</i>		1.000	.658
<i>Manyika J, 2011</i>		1.000	.647
<i>Mcafee A, 2012</i>		1.000	.639
<i>Mcafee A, 2014</i>		1.000	.839
<i>Nelson R, 1982</i>		1.000	.685
<i>Vasarhelyi M, 1991</i>		1.000	.571
<i>Waller M, 2013</i>		1.000	.758

*Extraction method: principal component analysis*

Source: Authors, 2015

<i>Component</i>	<b>Total variance explained</b>								
	<i>Initial Eigenvalues</i>			<i>Loading squared</i>			<i>Rotating arrows loading squared</i>		
	Total	% of variance	% cumulative	Total	% of Variance	% cumulative	Total	% variance	% cumulative
1.	4.611	27.126	27.126	4.611	27.126	27.126	3.480	20.470	20.470
2.	3.166	18.624	45.750	3.166	18.624	45.750	2.725	16.030	36.501
3.	2.296	13.506	59.256	2.296	13.506	59.256	2.424	14.261	50.761
4.	1.310	7.708	66.964	1.310	7.708	66.964	1.921	11.299	62.060

5.	1.070	6.293	73.257	1.070	6.293	73.257	1.903	11.197	73.257
6.	.977	5.745	79.002						
7.	.742	4.362	83.364						
8.	.579	3.405	86.769						
9.	.486	2.860	91.904						
10.	.387	2.275	93.651						
11.	.297	1.747	95.298						
12.	.280	1.647	96.693						
13.	.237	1.395	97.915						
14.	.208	1.222	98.775						
15.	.146	.860	99.546						
16.	.131	.771	100.000						
17.	.077	.454							

*Extraction method: principal component analysis*

Source: Authors, 2015

We found that there are five factors that explain 73% of the variance. Factors represent the group of authors who were cited for the same work and therefore are the main references for the texts that found from the used

keywords. In Table 3, the formed factors are presented, the authors that comprise each factor and their factor loadings.

**Table 3 – Rotated matrix with components and factorial charge**

	Component				
	1	2	3	4	5
Cukler K, 2013	.877				
Harrls J, 2007	.873				
George G, 2014	.828				
Manyka J, 2011	.562				
Lavalle S, 2011	.506				
Grant R, 1996		.859			
Kogut B, 1992		.783			
Nelson R, 1982		.758			
Baney J, 1991		.698			
Chen H, 2012			.831		
Mcafee A, 2014			.726		
Dean J, 2008			.674		
Waller M, 2013				.795	
Boyd D, 2012				.720	
Bolen J, 2012					.843
Vasarheyl M, 1991					.742
Mcafee A, 2012					.553

Source: Authors, 2015.

Verification of internal reliability of each factor was tested with verification of Alpha Crombach. For factors 1 and 2, which means 45% of the variance, Cronbach's alpha was greater than 0.60, as recommended by Hair (2014).

The factors 3.4 and 5 have a lower alpha Crombach than 0.60 therefore be highlighted that one of the limitations of the work. Although we may present the appointment of all the factors, we will focus more on factors 1 and 2, with internal reliability, as recommended by the literature on the AFE. With properly identified more relevant texts, we began to study, through content analysis, what were the similarities involved in the authors grouped the same factor, then going to name these factors.

## Factor analysis

**Factor 1**, formed by the authors Cukier K, 2013, Harris J, 2007 George G, 2014 Manyika J 2011 and 2013 and Lavalley S, 2011, (Cukier & Mayer-Schoenberger, 2013; Davenport & Harris, 2007; George, Haas, & Pentland, 2014; Lavalley 2011; James Manyika et al, 2011) consist of publications about the recommendations on how to search or to approach the topic Big Data and its potential value to the business and business in general. Are mostly technical publications arising from technology and consulting firms - mainly IBM, Accenture and McKinsey- (Ballard et al, 2014; "Big Data Research - Accenture," 2016 "Big Data," 2014; Jewell et al, 2014; Quintero et al, 2015), supporting the importance of considering the Big Data in the process of value creation and innovation.

Also underscore the importance of developing the analytical capacity of these databases - analytics - considering that it is the construction of a different capacity, so a capacity of differentiation and competition in the market. Even the recommendation of the editors, written by George G. In 2014, (George et al., 2014) is editor of the Academy of Management Journal, which aims to guide the possible lines of academic research theme, which recognizes the skills to handle issues related to Big Data in the companies, appear to represent a new distinctive competence and therefore has the potential to create value and innovation.

Because these are the relations between big data and skills that can be achieved by the business, the development of analytics skills, their ability to mobilize new skills and new capacities, generating competitive advantages, we call this factor "Data Capability".

By analyzing **the factor 2**, formed by the authors, Grant R, 1996 Kogut B 1992, Nelson R, 1982 Barney J, 1991 (Barney, 1991; Grant, 1996; Kogut & Zander, 1992; Nelson, 1982) found the theoretical currents involved in theme Big Data and innovation to date. Verified the presence of the theories of Dynamic Capabilities (Teece & Pisano, 1994),

Transaction cost theory (Williamson, 1981), Evolutionary Theory (Nelson & Winter, 2009) and Resource Based View (Barney, 1991), proposed by Vision based on knowledge (Grant, 1996; Kogut & Zander, 1992). The theoretical dominance, however, is related to the Resource Based View and principles dealing with the competitive advantage. Considering the Big Data a source of new skills and capabilities, the authors seem to consider that these skills could not only be a competitive advantage *per se*, but also enhance the value of existing resources in the company, making it the rarest, most inimitable and bravest. In this way of thinking, skills and capabilities arising from analysis of Big Data, seem to adjust to the theoretical principles of both the Resource-Based View, as the Dynamic Capabilities. Because it is predominantly theoretical work, which relate the theoretical basis for building competitive advantage, differentiation and value through Big Data, called this factor "Data Value Theory".

**Factor 3** is formed by the authors, Chen H, 2012, McAfee A, 2014 Dean J, 2008 (Brynjolfsson & McAfee, 2014; Chen, Chiang, & Storey, 2012a; Dean & Ghemawat, 2008), are predominantly empirical - except Chen H, 2012 which is descriptive-seeking exemplify and argue, as big data can be used, relating a goal and a way to use, offering use of recommendations related to the processing and storage of data. Are more operational studies, with contributions to the substantive knowledge and therefore call "Big Data Purpose".

**Factor 4**, formed by the authors, Waller M, 2013, Boyd D 2012 (Rawlings, Waller, Barrett, & Bateman, 2013; Boyd & Crawford, 2012) discusses the technical use of the database to solve organizational problems. Both are a reflection of possible uses Big Data to better understand a culture or population a technological propensity or even an important tool for academic research, encouraging the appointment of this factor as "Use Big Data."

The **5 found factor** is formed by the authors McAfee A., 2012 and M. Vasarhelyi, 1991 (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012; Vasarhelyi & Halper, 1991) focus on which approach would be most effective for draw results through Big Data. This theme led us to define this factor as "Data Mining Methodology of Big Data."

## Literature review

The significant increase in capacity creation and storage of data, made possible by the start of the digital era in 2000, was one of the drivers of the term Big Data creation (Hilbert & Lopez, 2011). Currently the term Big Data definition considers that this term is information assets, characterized by a high volume, where the speed

and range should require a specific technology and analytical methods for processing value (Andrea De Mauro, 2014). The term originally arose in the META Group 2001 report (Gartner company) by the analyst Doug Laney, which defines the data growth challenges and opportunities as being three dimensional, ie the increase in volume - the amount of data - -speed data rate into and out -, and -gamma varieties or types of data sources and data-(Laney, 2001).

Although the report has shown the dimensions that characterize the Big Data, researchers continued to seek a definition that best explain this phenomenon. One of the most used settings for Big Data is precisely defined by the Gartner Group, who understands as assets formed by high-volume information, speed and variety. These characteristics require debates about the cost-benefit involved in its mining, for the purpose of creating value and innovation in various businesses. Innovative ways of processing information for greater visibility and decision-making also become concerns of the researchers of this issue, as we can see in Beyer (2011).

The definition of the IDC - International Data Corporation, is also widely used, however, it presents more instrumental and operational form of conceptual that "... the Big Data technologies describe a new generation of technologies and architectures designed to extract economically the value of very large volume and a variety of data, enabling high speed capture, discovery, and / or analysis "(Hilbert & Lopez, 2011). This concept was ratified in the work Taurion called Big Data (Taurion, 2013).

The consensus among researchers of the subject, is that the use of Big Data turns to his analysis, but this analysis is much more powerful than the analysis of the past (McAfee & Brynjolfsson, 2012). The author explains that today useful information for organizations may come from social networks, images, sensors, web and other unstructured sources. This tends to significantly increase management challenges, requiring decision makers to learn to ask the right thatstions to take their evidence-based decision. To be able to analyze and extract the data organizations need to hire scientists can find patterns in large data sets and translate them into useful information for their business (McAfee & Brynjolfsson, 2012).

More than performing intersections, find commonalities, distances or patterns in the data, the generation of value and innovation need to be considered approaches and further appropriation of the meaning of the data, which related literature will be verified in the next section.

### **Big Data and value generation**

The Big Data can contribute greatly to the future of corporations, who must decide their processes from

quality of information in the shortest possible time, ensuring competitiveness within which they operate (Volpato, Rufino, & Day, 2014): "It is of fundamental importance the decision maker to have at its disposal the most relevant and useful information, allowing to minimize the subjective effects and increase the influence of reason on the outcome of the process. Faced with a large volume of variable data, which are distributed quickly, it is necessary for the firm to ensure the veracity and the value thereof. From a data integration system, you can organize them, categorize them and filter them to ensure consistency of information "(Hilbert & Lopez, 2011).

The Big Data bases of data analysis refers directly to the theories on Business Intelligence (BI) & Analytics, and its technologies are based mainly on data mining and statistical analysis. Most search technithats of deep data has commercial technologies to perform this procedure (Chaudhuri, Dayal, & Narasayya, 2011). Since the late 1980s, several data mining algorithms have been developed by researchers in artificial intelligence, organized in communities that discuss and develop solutions for databases. In 2006 IEEE International Conference on Data Mining (ICDM), the most influential data mining algorithms were identified based on expert appointments, citation counts and the research community.

These algorithms seek to cover the classification, clustering, regression, association analysis and network analysis. Most of these algorithms for mining of popular data were incorporated into the commercial environment and on open source data mining systems, democratizing somehow access (Clemmensen, Hastie, Witten, & Ersbøll, 2011).

Business intelligence and analytics -BI & A- and their fields of large data analysis have become increasingly important in both communities, academic and business over the past two decades. For example, based on a survey of more than 4,000 information technology (IT) professionals, in 93 countries, 25 different industries, IBM Tech Trends Report identified in 2011 the business analysis as one of four major technology trends in the 2010s (Chen, Chiang, & Storey, 2012b). In a business analysis of the state of the survey, conducted by Bloomberg Businessweek, also in 2011, it was found that 97% of companies with revenues over \$ 100 million have been identified as users of some form of in-depth analysis of business ( Chen et al., 2012b).

The McKinsey Global Institute report (J. Manyika et al., 2013) predicted that by 2018, the United States will face a change of 140,000 to 190,000 people with deep analytical skills, but there will be a 1.5 deficit millions of database managers with the ability to analyze big data to make effective decisions.

Chen et al. (2011) warns that Hal Varian, chief economist at Google and professor emeritus at the University of California, Berkeley, said in 2011 on the

emerging opportunities for IT professionals and students in analysis of the data as follows: "So, that is getting ubiquitous and cheap? Data. And that is complementary to the data? Analysis.

So my recommendation is to have lots of courses on how to handle and analyze data: Databases, econometrics, statistics, visualization, and so on "The opportunities associated with the data and analysis in different organizations have helped to generate significant interest BI & a, which is often referred to as the technical, technological, systems, practices, methodologies and applications used to analyze critical business data to help a company better understand your business and your market, doing business with agility and making more consistent decisions. In addition to the underlying data processing and analytical technologies, BI & A includes business-centric practices and methodologies that can be used for various high-impact applications such as e-commerce, intelligence market, e-government, health and safety (Chen et al., 2012a).

Through the initiatives of BI & A 1.0, companies and organizations from all sectors began to gain a critical view of structured data collected through various enterprise systems and analyzed by database management related to business systems. Over the past few years the web intelligence, web analytics, Web 2.0 and content capacity generated by the user, not normally structured, marked the beginning of a new era of BI & A, which generated research 2.0, leading to intelligence unprecedented on consumer opinion, customers' needs and recognizing new business opportunities.

While we are writing this study, this time analysis of Big Data, even as the BI & A 2.0 is still maturing, we are balanced on the edge of BI & A 3.0, with all the uncertainty stemming from new technologies, potentially revolutionary and all the advantages and problems they may bring us.

The investigation of these data, hitherto invisible to ordinary users, is intended to serve, in part, as a platform and conversational guide to examine how the analytical discipline can better meet the needs of business decision makers needs, considering the maturity of technologies emerging BI & a, the omnipresence of Big data and failures from data managers, even when experienced, coupled with the lack of professionals with deep analytical skills, beyond the limited rationality inherent in our human nature.

The demands proposed by Big Data, not only affect practitioners, as well as technology and business schools. You may need to create a new vision for the Information Systems programs to address these and other issues that are permeating the new reality of the future Information Technology professionals.

The creation of value and innovation for business starting from a correct analysis of arising database of a Big

Data may be important for various applications in today's society, such as e-commerce, business intelligence, e-government, health, public safety and security in general. When mapping data samples, using the current BI & A, you can generate a knowledge of the entire population of data, hoping to thus contribute to decision making, create - cognitive bases - knowledge sources and increase the current discussions the importance of scientific research - relevant - on Big Data. Examples of how big data analysis can be applied, are found in the article by Melo et al. (2014) called Big Data Is the Next Big Thing in Performance Measurement Systems?

According to the authors, this analysis can be applied to continuous process monitoring to detect situations such as changes in consumer sentiment; to explore the network of relationships, as friends suggested on Linked In and Facebook; to identify a real-time fraud; to understand why the defect rates of automakers surged; to continuously scan and intervene in health care practices; and to better anticipate the online sales based on a set of data characteristics of a particular product (Demirkan & Delen, 2013).

Such perceptions seem to indicate an interesting relationship between Big Data and PMSS. Tien (2013) proposes a different approach to the analysis of Big Data, and this will be the basis for customization and can generate the third industrial revolution (IRR), according to the author. IRR is based on the confluence of large data analysis, adaptive services, and digital manufacturing, centering on the integration of services or products and would be starting in the second decade of the 21st century.

Investments in Big Data point to this fact, there is seen the amounts involved for the development of Big Data as a "field" or "field".

In 2012, Boston University received \$ 15 million to create the Institute Rafik Hariri B. Information Technology and Computer Science and Engineering, an interdisciplinary research center for discovery through the use of computational approaches and data-driven and advances in computer science (Chen et al., 2012a).

The Big Data could be an alternative to the failures of traditional PMSS, because only when we define the data, information and analysis we can see that the traditional measurement mechanisms do not work efficiently. Organizations should be concerned with the analysis of service and accuracy, involving the quality, and the cost and delivery time of these analyzes (Demirkan & Delen, 2013).

Organizations still need to make changes in five areas to reap the real benefits of Big Data. They need to change the leadership, which should set clear goals related to the use of data in the routines and decisions by providing performance indicators that reflect the success sought by the organization, considering the human aspects,



therefore the data are very important, but insight human is key.

Talent management is also important, it is imperative to hire the right people - driven data -. It is necessary to invest in new technologies, including appropriate software.

Finally, the company culture should reinforce the use of data in decision making more than the managers themselves (Waller & Fawcett, 2013). Proper operation of Big Data can help improve organizational performance, but first it is necessary to transform the corporate culture of organizations and their capabilities. Efforts should be made to weave the Big Data into the fabric of daily operations (Waller & Fawcett, 2013).

Some benefits of using big data are: better integration and analysis of quantitative and qualitative data; More accurate predictions (Waller & Fawcett, 2013); and making processes more efficient and effective decisions, which make more agile and efficient business (Demirkan & Delen, 2013). Companies that are at the top in terms of financial performance, - their industry - and use making resource decisions based on data on average 5% more productive and 6% more profitable than their competitors who do not use this feature (McAfee & Brynjolfsson, 2012).

## **Big Data, information, innovation and value**

Data are symbols that represent certain circumstances, phenomena, facts, whether physical, chemical or behavioral. By assigning specific order, these codes symbols has the information discussing here is not only the formal aspect of the information, but the broader sense, ie, collect, organize, analyze and interpret the information. According Bregonje (2005), when information is applied, it becomes aware. You can then make the analogy that information applied solution of the actions of a problem, produce the discovery or invention, these in turn, Innovation background. From the evidence of Bregonje (2005), one can infer that the information can moderate the relationship between basic research and scientific discovery or invention, as can moderate the relationship between applied research and invention and ultimately moderate relations between invention and innovation.

Considering the challenges for the treatment of big data, we can understand the processes involved in its storage and treatment, also known by the English term, Analytics, has the potential to contribute to innovation, the very definition of innovation. Innovation is the availability of something substantially new in varying degrees of novelty or type that generates economic or financial value (OECD & Eurostat, 2005). When Bregonje (2005) argues that the applied information becomes knowledge, we can understand that Big Data stores data when encoded,

transformed into information and that information, when applied, are transformed into knowledge.

Knowledge is the element that adds value, that is, promoting the value proposition in business models (Osterwalder, 2009; Enzmann & Schomer, 2013; Flower & Maglio, 2004). Whereas something can only be considered innovative when it generates economic or financial value and that knowledge is the element that promotes business value proposition, we can understand that there is a relationship between knowledge element and value to the business. Innovation is in this case, intrinsic to the process of construction of knowledge, and this in itself knowledge.

A reverse reasoning, think of knowledge as the application of information and that information from the encoding of the data leads us to understand the connection between the Big Data Innovation and Value for business. Obviously, we need consider only the theoretical thread as a proposition, saved all the difficulties, obstacles and challenges, especially posed by Buhl et al. (2013). Some of the challenges are being faced by both academics as by practitioners of innovation based on Big Data. Some of the cases will be discussed in the following sections such as the structure of the data, the data recovery process (Crawling and Data Mining), the processing of data (KDD) and some common forms of analysis.

## **Structured and non structured information**

In the universe of data when stored in some repositories, there are two large groups, called structured data and unstructured data. It is understood by those structured data that are expected or known in their way since the entry, storage and output normally being complete data. They are typically stored in databases that use SQL (Structured Query Language). Its main features predictability of how data will be stored and enter the database, helping to create search indexes, which results in higher performance in data recovery when needed. The security aspect is also increased because the relational structure.

Adding a random data, either for attempted fraud or systemic error in relational data structures are not simple tasks. Unstructured data are those that may originate from different sources, in different formats, the entry can be carried out completely or incompletely.

The databases are called NON-SQL, as databases that do not use the SQL structured linguagem. Are databases that allow data storage that principle were not expected, as opinions, feelings declared in social networking, data from sensors and other machines. Security issues can be identified as the trade off to be thought. Since no relational structure, the insertion of data which are not part of the storage environment may occur more easily.

## Big data main analysis process

The literature shows several ways to access and enjoy the Big Data, usually in competitive intelligence literature, however, big data analysis processes are virtually countless, because depend on each database or combination of databases, the context and interest of analysis, continuity or perecividade data, of economic resources, technological and human disposition, described by Fleming et al. (2001), Machado (2004) and Wu et al. (2014) (Fleming & Sorenson, 2001; Machado, 2004; Xindong Wu, Xingquan Zhu Gong-Qing Wu, & Wei Ding, 2014). There are steps or common sub-processes in the data analysis process, described in the literature reviewed, which will be discussed briefly in the next section.

## Crawling, Mining, Analyse (KDD)

Put simply we can say that to harness the potential of Big Data, we need to extract data (Crawling), separate data of interest (Data Mining) and analyze these data (Analyse). If we call these steps sub processes, the combination of which forms the process called knowledge discovery in databases (Knowledge Discovery in Data Base - KDD), (Tian Zhiping, & Zhengyin, 2013).

The stage of extraction or data recovery, is to identify the most appropriate way to recover the data that we want to avoid noise.

Noises are extracted data together with the mass of data retrieved, but in fact not part of the data you seek. The noise may not be totally eliminated but should be minimized and therefore the crawlers comprise data validation algorithms. Such algorithms are those who try to answer the search expressions. Search expressions are words or terms used as input a search. He popularized the use of Google, and most people use words and terms to look for answers in the Google search engine. The words and terms typed into Google to get specific response, are expressions of search and crawling is done by Google itself that presents the recovered results for that search phrase on the screen.

Data mining or data mining is the separation procedure of the recovered data into useful and useless data for that search term. At this stage, there is still some noise, the crawler algorithms not treated and also content that meet the search expression, but does not represent the expected ex ante results. This fact presents itself as one of the challenges mentioned by Buhl (2013), the truth.

Truthfulness in this study is related to the identity of the data recovered, the decision was, in fact, it belongs to the context declared by the search expression. The fields of study that deals with the Semantic Web, have concerns like that, seeking to contribute both in the sub process of data mining, and in the sub process analyse.

The sub analyze process can be defined as the iterations involving data coding to be transformed into information. In the case of iterative process, we can not define, observing the literature, its extension in volume of data analyzed or its speed of analysis, since it seems to be a non-linear sub process (Martínez-Román, Gamero, & Tamayo, 2011).

The iterations required, is not linear are more related theories of complex systems. The fact that from the complex to define simplified models with significant degrees of all the explanation for after this involved again complexity, iterating this sense leads us to think of a process in a spiral (Fleming & Sorenson, 2001; Prior, 2013).

The spiral, shows that every iteration, although in the same horizontal pass, there is change in the vertical position and vice versa. Another way proposed in the literature to design complex systems models of explanation is in network form. Networks may have non-linear forms and represent some dimension of complexity that aims to understand. In a network, the direction and intensity of the elements or nodes (nodes) are constituted are not predictable (Yugue & Maximiano, 2013).

Thus, to analyze large volumes of data, non-linearly, not structured or connected with iterative processes show other challenges posed by Buhl (2013), regarding volume, speed and range. According to sustained here, the sub process of data analysis is what encode, turning them into information.

Therefore, methods used in scientific research, specifically the quality and quantity can contribute to this sub process of KDD and therefore will be presented in the next section, in the context of the analysis of Big Data.

## METHODOLOGICAL PROCEDURES

As this is a new proposal for the definition of Big Data, mixed methods were selected, involving bibliometrics and content analysis. Bibliometry was developed to raise the main literature, worth to both the database Web of Science, whose procedures were detailed in bibliometrics section in this article. The content analysis was used to recognize patterns and categories as recommended by Bardin (1999) and we selected three of the main scientific articles found about Big Data, published by MIT Sloan Review, providing the foundation of our conceptual definition for Big Data.

The content analysis we are proposing, is divided into quantitative analysis involving the frethatncies of occurrence of words in the set of extracted texts of selected scientific articles (corpus), combined with confirmatory factor analysis and Reinert grouping method (1998) .

The second part of the analysis is qualitative and was held on the results of content analysis, resulting in the

mapping of ontology found on Big Data. After that brief explanation of the content of this section, we will detail each of the steps outlined in the following subsections.

The texts chosen to perform the content analysis have been identified in bibliometrics, preferring items they were younger and with higher quotes. In addition, smaller specialized cores were identified as the concentration of authors, outside the areas of concentration described in Figure 1, but although publish more isolation, feature 7080 together citations distributed in 20 articles.

This is the case of MIT Sloan Review, and the articles that met criteria were established as How 'big data' is different from TH Davenport, P Barth, R Bean (2013) 96 citations, Big Data, analytics and insights from the path to value S LaValle, and Lesser, R Shockley (2013) with 208 citations and Strategic outsourcing: leveraging knowledge capabilities of JB Quinn, ES Strategy (2013) with 2125 citations.

### **Quantitative analysis**

Depending on the nature of the research and data available, may be more contributory to use quantitative methods. Usually when there are answers to questionnaires or more structured information, the quantitative analysis can be complementary to the data analysis process. The most common tests are the average tests and hypotheses as bivariate parametric tests and their derivations as regressions. Multivariate data are typically used the factor analysis, cluster analysis and structural equation. The choice of test depends on the nature of the search expression and data available, always considering the minimum sample sizes and other parameters defined quantitative method used.

These methods of analysis have instrumental character, aiming to apply the information to obtain knowledge and value. Competitive strategies, supported by competitive intelligence, you can receive the Big Data context of contributions, and these aspects are presented in the following section.

### **Reinert method and factorial analysis of correspondence (AFC) – Text Mining**

In any language there is a lexical representation, which is the set of codes used to represent the "signs" or meanings of words, allowing communication to occur between people. Verbal communication uses the sounds, in the case of speech and also uses codes in the case of written communication.

The Big Data need not be formed only by verbal communication, such as sounds and written data, but also with non-verbal communication such as images and

symbols, which are not addressed in this study. The focus will be on the written verbal communications.

Content analysis is the qualitative method used to identify the essence of a verbal information written corpus. This method can be divided into frequency analysis, hierarchical classification of Reinert and Confirmatory Factor Analysis not exhaustively. For treatment of verbal language writing, we can use the method Reinert (1989), which highlights the lexical proximity between the words, producing a map of classes of words that make up particular database texts. These texts may be interviews on consumer reviews, complaints, suggestions and others. Analyze 30 or 40 suggestions could be done manually, but analyze 10,000, 100,000 or 1 million suggestions, it is usually not possible manually.

In such cases, applying the method Reinert (1989), automated through software, can allow analyze 1 million suggestions of texts or customer complaints in a few hours.

In this same line of reasoning, evaluating qualitative information, we can perform the confirmatory factor analysis (CFA) considering the lexical distance of words. We realize that we are using a quantitative technique AFC, but the nature of the research is to deepen in unstructured data belonging to a particular context, no interest in extrapolating or inference. From the AFC, they are constructed factors and factor loadings of the set of related words.

The AFC separates the words by lexical proximity, showing the analyst, which the themes present in the database and how they relate. The interpretation of such data, aided by the mentioned methods will give the data encoding and the meaning necessary to pass to be called, useful information for application and value creation.

### **Qualitative analysis**

When we need to know in more depth on particular research topic, you can use the qualitative analysis. Identify the reasons and justifications for the occurrence of certain phenomena represented in the data, can lead to ideas about how to make decisions on which products to invest, services to be developed, processes for offer or improved organizational arrangements and possible plans marketing to build.

This study aims to provide only a snapshot of the overview of the universe of obtaining value from the big data analysis, so details on qualitative and quantitative methodologies will not be developed, just presenting the information regarded as relevant for the introduction of such methods in the context of the analysis of Big Data.

## Strategy of innovation using Big Data and competitive intelligence

The authors Demirkan and Delen (2013) seek guidance on some strategies that organizations can use to gain competitive advantage through the use of Big Data. His statement refers to the use of decision support systems oriented services (DOS cloud). According to the authors, this is one of the key trends for many organizations, hoping to become more agile. For this to happen, they propose a conceptual framework for DOS. This sets the perspective on how to align the DOS environment-oriented product, and demonstrate the opportunities and challenges of engineering-oriented DOS cloud services, such as a cycle that feeds back.

When defining the data, information and analysis as services, it was realized that the traditional measurement mechanisms, which are mainly the time and cost, do not work well. So organizations need to consider the value of service and quality level, and the cost and duration of services.

The cloud DOS allows the development of scale, scope and speed, generating savings for organizations, tying the prospects of information technology strategy with the database perspective, creating a continuous innovation strategy processes within organizations (Demirkan and Delen, 2013).

It can also analyze the role of decision environment to explain how the business intelligence capabilities (BI) are utilized as a strategy to achieve success. His procedure was

to analyze the decision environment in terms of the types of decisions taken and the organization's needs, as the processing of information, supported the best results of the previous decisions (Işık, Jones, & Sidorova, 2013).

The authors contribute with the Big Data field of study, proposing that the technological capabilities, such as data quality, user access and the integration of BI and other systems are necessary for the success of BI, regardless of the decision environment, but the decision environment influenced the relationship between success and the dynamic capabilities of the organization, such as the extent to which BI endured, flexibility and risk in decision-making when it became in fact, an organization's strategy (Işık et al., 2013).

Because the purpose of this article, descriptive analyzes were used to identify the categories of most frequent words and the factor loadings of the word groupings found in articles, to compare with the proposition of the business model ontology Osterwalder (2009). From that mechanism, it was possible to compare the patterns present in the scientific articles dealing with Big Data and possible contributions to the business model of ontology Osterwalder (2009).

## FINDING AND RESULTS

By analyzing the three texts on Big Data, held the Reinert analysis and confirmatory factor analysis in relation to the ranking of classified words. The factor loadings found are described in **Table 4**.

Table 4: Factorial charges of factorial analysis of correspondence

	<i>Factorial charge</i>	<i>%</i>
<i>Factor 1</i>	0,35192	30,16
<i>Factor 2</i>	0,31543	27,03
<i>Factor 3</i>	0,2784	23,86
<i>Factor 4</i>	0,22094	18,93

Source: Authors, based in Iramuteq software, 2015.

The network formed by the analysis of three articles demonstrates the presence of the words that showed a significant level of similarity, as the factor loadings presented in Table 2, highlighting the words best placed in the ranking. The words most frequently and with their

respective groups on factors that will be discussed in this article due to be apparently more relevant to the generation of value and innovation for business are the words: business, process and organization. Figure 2 shows the network formed by analyzing the factors described in **Table 1 in graphical form**.



Buhl (2013) are unable to minimize the problem of conceptual obscurity on the topic Big Data and only expanded the operational vision hitherto three-dimensional for multidimensional view. We are insisting to know the dimensions of the theme or even its elements does not guarantee your understanding of the phenomenon.

Considering only more superficial and shallow these definitions in scientific terms, we can understand, because there are Suspicions about the relevance of Big Data as a scientific subject and speculation that It may be a passing fad. In one hand, we have the major information technology vendors for Big Data, Advocating the issue, which may refer the technical and scientific community to identify the conflict of interest. On the other hand, we have the technical and scientific literature, only operational definitions, functional and tooling, Which Ultimately Prevent deepening the theme, to be able to study actually it as a phenomenon, from the point of view of social sciences applied.

We found that the Resource Based View - RBV - is the present theory in scientific publications about Big Data, Which Allows us to approach the Big Data resource characteristics described by Barney (1991), is competitive advantage of reach. Resources are rare, valuable, imperfectly imitability and without equivalent (Barney, 1991). The authors have located in our bibliometrics agree that Big Data is and will be used as a source of competitive advantage and this may be the link that we were looking for value creation and innovation in business. We argue here that the data transformation into information and knowledge into then, as described by Bregonje (2005), do not guarantee the creation of value and innovation for business and for businesses.

When the knowledge generated from Big Data is applied to processes, products and services, and Thus the generation of value and innovation will not be automatically guaranteed and that uncertainty, in our view, is what to Researchers, theoretical and empirical BOTH, it had not We have been exceeded in the articles we reviewed. In order to contribute to this discussion, we find the dominant theory in our bibliometrics, which is the theory of Resource Based View. The identification of the dominant theory in scientific publications on Big Data, lead us to identify there is evidence of value creation and innovation in business and companies, can be through Reached the Big Date. If the knowledge derived from the analysis of your information -codification data - may looks like, in terms of features, with the features described by Jim Barney (1991) - rare, valuable, imperfect imitability and without equivalents.

**Rare resource;** the rarity, it seems natural that each set of data that can be by captured the company will be unithat, because When the date Referred to in its own

transactions are combined - structured data such as sales and customer complaints received in the central relationship - and the data refer to the behavior of Their social interest groups - unstructured data, such as coming from networks social forms-something unithat. The chance of two companies combine structured and unstructured data the way same and subsethatntly performing the same analysis can be Considered minimal and perhaps nonexistent.

Obviously, we know that something unithat Is not Necessarily and automatically rare considering the rarity something that is beyond exclusivity, However, be only increase the chances that something is or Become rare.

**Valuable resource;** the applied knowledge, originated in the Big Data are valuable, If They allow support making evidence-based Decisions - data driven - enabling make more efficient and effective company, Minimizing errors, losses, waste, directing what should not be produced, When It should not be produced and who should not be delivered.

**Resource imperfect imitability;** about the imperfect imitability, it is Necessary to note that, the Barney described (1991), even if knowledge is rare, it may be the target of an attempt to copy, innovative strategies, However, to generate competitive advantage, value and innovation, the imitability can not be perfect. In our research, the effects of analytics technic thats and human insights were argued the essential to the use of Big Data and beyond agree with the argument, added that such procedures can minimize imitation possibilities of knowledge originated in Big Data. We want to include the organizational culture built with guidance on the use of data and evidence, can greatly hinder imitation, Therefore, the knowledge derived from Big Data will not be inimitable *per se*, but become more and more inimitable the they are improved the analysis of knowledge - called analytics commercially and scientifically known the Science- data, learning of social actors Involved in data usage process - human insights - and culture-oriented company - driven-data. What we are saying, in other words is that when we coordinate what we found in our studies described above and RBV, we infer that use of Big Data needs to be the culture of companies present in Their routines of everyday life and not a specific or department sector for two main Reasons: a) If the Big Data analysis capability is dammed in a sector or a few people, this capability can be located, harassed and imitated with high chances of mobility to a competitor and, b) the difficulty of imitating something non-clustered and that is part of people's behavior, it makes bigger, specially because each person specializes and appropriates a part of knowledge, providing results in the form of synergy, like a 'knowledge of fabric, "which

makes sense only when it acts together matching resources;

Finally, the knowledge by generated the analysis of Big Data **has not yet equivalent** even discussions on the use of feeling and experience, they become vulnerable due to limited rationality inherent to human beings, the dynamism of the. Data sources and their growth rates, as well as the speed with which change habits and social behaviors. Associated with great speed with which societies have changed, we still have colleagues such changes are not immediately perceived when change begins, however, it takes up even years to realize that there are other values and needs in society.

The principles proposed by this study, whose understanding may influence the generation of value and innovation of the companies, are based on knowledge originated in the Big Data in the resources previously sustained. From the analysis of the characteristics of knowledge as a resource, we are proposing the principles of the phenomenon, which will lead to the desired theoretical conceptualization. The rarity of knowledge originated in Big Data is related to the heterogeneous nature and pretty much "chaotic" the types and sources of data, which arise from the human cognition. There seems to be a one data even that has not been derived from the action of the agent social - the purchase transaction, sale, the post on a social network, a text message etc ... - or the agent cybernetic programmed by social worker - internet case things Involving sensors, electronic devices, appliances etc...-. These features, cut out the principle of "cognitive representation" and the principle of "diversity" another identified principle is related to the shape and location of the date. These are distributed data and Therefore "ungrouped". They consist of structured and unstructured data, so in the "nature" will be found ungrouped. At this point, it clarifying our understanding, based on studies that starts from the moment the data is grouped - even on databases that let you integrate structured and unstructured data - the database is no longer the Big Data and becomes a database, which will be used for the entire process of analysis and construction of knowledge. We argue, therefore, that to create value and innovation from Big Data, you must choose which date will be integrated into the analytical basis. Seems to get clearer than using the lens of RBV, the chances of creating value and innovation will be higher. The third principle identified, is related to dynamic data. This is a phenomenon that arises from data capture to "dynamic incremental." This means that is both the variety, the amount of the data varies, increasing, due to the economic social or exchanges.

This fact has the effect of geometric growth, the more people transacting, and more transaction technologies are made available, the more data is generated, therefore, the

speed at which date is generated in terms of megabytes, for example, Also Increases. We are arguing that the three dimensions described in the literature - variety, volume and velocity is the effect produced by the principle of "dynamic incremental."

Listed the principles, we believe it is possible to consolidate the conceptual proposed definition, based on principles that are anchored part in RBV and part of the content analysis on the most cited articles found from our bibliometrics: Big Data is a cognitive social and cybernetic representation relations, consisting of heterogeneous data sources and characteristics, ungrouped and incremental presenting dynamic in terms of quantity, variety and speed.

Considering the principles and the definition that we are holding this paper, we carry out the survey of some propositions related to the topic, considering some practical contributions. The Big Data has the one of its principles "cognitive representation." We are insisting that this principle comprising an analyst, you can check what are the behaviors performed in a period of time by a particular social group and compare it with social groups of different characteristics - demographics, psychographic, behavioral - checking variations in time. This allows us to understand in which direction society is going and anticipate possible needs. Despite the widespread respect of the term Big Data with BI - Business Intelligence acronym, we offer the first proposition of this study, aiming to generate value and innovation to business: 1. The initiative towards the use of Big Date is assignment of managers, rather than information technology or other areas and this order has the greatest potential for value creation and innovation.

This proposition is due to the use of knowledge as a resource, and other words, since that Big Data is directly related to decision-making at various levels of management, allocation to understand the importance and endeavor to use it, makes the process of choice, acquisition and use, more efficient and assertive, with the greatest potential to create value and innovation for business.

On the other hand, the use of Big Data is suggested or imposed by particular area of business, whether the own area of innovation or the area of information technology, potential this is less or non-existent, since it will be the tool more than management teams will be required to use, without interest indeed. This is due to colleagues all the information.

That is acquired in a BI analysis form the basis for decision-making in organizations or even be part of the business model of the company and such decision activities, follow beliefs and values manager. If the knowledge generated by Big Data is not part of this set of values, these managers will only make the minimum

required in the use of Big Data, which certainly undermine the value creation and innovation.

In our content analysis, the words "analytics" and "information" were expected due to the constant presence in texts on Big Data, however, the presence of "process" and "organization", present the hitherto words unexpected but provide new insights, about future theme holdings. From bibliometrics performed, we found the incidence of classical authors of organizational strategic studies. This point alone security to offer in the second proposition:

2. The results arising from analysis of Big Data influence one strategy of companies by both the Resource-Based View, the development of the new capabilities - Dynamic Capability.

The list of factors showed by our bibliometrics indicate two more proposals, which will not be discussed here, but that can be developed in future work, as follows:

3. In the RBV, Big Data may reveal new ways of transacting in social and economic terms, to point to base the design of an innovative business model.

4. The study of Big Data Relates to the new thinking (insights) organizational innovation, since it is the representation of the most current thinking of the social group or the society.

These propositions begin to bring the Big Data phenomenon to the reality of social and applied sciences, management and administration, as well as to report it to the characteristic constructs of business management, considering the knowledge generated the resources that can generate value and innovation for businesses.

## REFERENCES

Andrea De Mauro, M. G. (2014). What is Big Data? A Consensual Definition and a Review of Key Research Topics. <http://doi.org/10.13140/2.1.2341.5048>

Baaziz, A., & Quoniam, L. (2013). How to use Big Data technologies to optimize operations in Upstream Petroleum Industry. *Journal IJI* (e-ISSN:2318-9975), 1(1), 19–25. <http://doi.org/10.5585/iji.v1i1.4>

## CONCLUSION

More than a new proposal for the definition of Big Data, we had the pretension to make clearer the principles involved in the Big Data as a phenomenon. We rely on studies, bibliometric and content analysis, through which we isolate three principles - cognitive representation, heterogeneity and ungroup and incremental dynamic - to support our conceptual definition: Big Data is a cognitive representation of the social and cybernetic relations, consists of data sources and heterogeneous characteristics, ungrouped and presenting incremental dynamic in terms of quantity, variety and speed. The propositions offered are communication possibilities of the theoretical elements discussed in the article and management practice, with main objective to create value and innovation for business.

We found that, in the scientific literature, business strategy, as well as studies on innovation seem to be approaching the Big Data phenomenon, timidly and future research should follow this trend. Our contribution in this article, is certainly minimal, however, we not had found until then a conceptual definition of Big Data, nor the more detailed description of their contribution to value creation and innovation. We argue that for Big Data to create value and innovation, the knowledge generated should be seen as resources, rare, valuable, imperfectly imitability and without equivalent, based on the RBV, to be the theory most used by authors located in our bibliometric study.

It seems therefore not treat yourself to a fad, but a latent phenomenon, which evolves in size, depending on variety and speed of evolution of social and economic exchanges that can be addressed by the applied social sciences because they are organizations and active participants in the formation of companies qua data strings make up the Big data and need sources of value and innovation. The field is proposed and looks very fertile.

Ballard, C., Compert, C., Jesionowski, T., Milman, I., Plants, B., Rosen, B., ... Redbooks, I. B. M. (2014). *Information Governance Principles and Practices for a Big Data Landscape*. IBM Redbooks.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17(1), 99–120. <http://doi.org/10.1177/014920639101700108>

Beyer, M. (2011). Gartner says solving “big data” challenge involves more than just managing volumes of data. Gartner. Archived from the Original on, 10.



Big Data Research - Accenture. (2016). Retrieved January 12, 2016, from <https://www.accenture.com/us-en/insight-big-data-research.aspx>

Big data: The next frontier for innovation, competition, and productivity | McKinsey & Company. (2014). Retrieved December 16, 2015, from [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)

Boyd, danah, & Crawford, K. (2012). Critical Thatstions for Big Data. *Information, Communication & Society*, 15(5), 662–679. <http://doi.org/10.1080/1369118X.2012.678878>

Bregonje, M. (2005). Patents: A unithat source for scientific technical information in chemistry related industry? *World Patent Information*, 27(4), 309–315.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company.

Buhl, P. D. H. U., Röglinger, D. M., Moser, D.-K. F., & Heidemann, D. J. (2013). Big Data. *Business & Information Systems Engineering*, 5(2), 65–69. <http://doi.org/10.1007/s12599-013-0249-5>

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88–98.

Chen, H., Chiang, R. H., & Storey, V. C. (2012a). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188.

Chen, H., Chiang, R. H., & Storey, V. C. (2012b). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188.

Clemmensen, L., Hastie, T., Witten, D., & Ersbøll, B. (2011). Sparse discriminant analysis. *Technometrics*, 53(4).

Cukier, K., & Mayer-Schoenberger, V. (2013). Rise of Big Data: How it's Changing the Way We Think about the World, *The Foreign Affairs*, 92, 28.

Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business Press.

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. *Commun. ACM*, 51(1), 107–113. <http://doi.org/10.1145/1327452.1327492>

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, 55(1), 412–421.

Enzmann, D. R., & Schomer, D. F. (2013). Analysis of Radiology Business Models. *Journal of the American College of Radiology*, 10(3), 175–180. <http://doi.org/10.1016/j.jacr.2012.09.001>

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research Policy*, 30(7), 1019–1039. [http://doi.org/10.1016/S0048-7333\(00\)00135-9](http://doi.org/10.1016/S0048-7333(00)00135-9)

Flor, N. V., & Maglio, P. P. (2004). Modeling business representational activity online: A case study of a customer-centered business. *Knowledge-Based Systems*, 17(1), 39–56. <http://doi.org/10.1016/j.knosys.2003.08.011>

George, G., Haas, M. R., & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, 57(2), 321–326. <http://doi.org/10.5465/amj.2014.4002>

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(S2), 109–122. <http://doi.org/10.1002/smj.4250171110>

Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65. <http://doi.org/10.1126/science.1200970>

Işık, Ö., Jones, M. C., & Sidorova, A. (2013). Business intelligence success: The roles of BI capabilities and decision environments. *Information*

- & Management, 50(1), 13–23. <http://doi.org/10.1016/j.im.2012.12.001>
- Jewell, D., Barros, R. D., Diederichs, S., Duijvestijn, L. M., Hammersley, M., Hazra, A., ... Redbooks, I. B. M. (2014). Performance and Capacity Implications for Big Data. IBM Redbooks.
- Joseph, J. H., Bush, R., & Ortinau, D. (2008). Marketing Research. McGraw-Hill Companies, Incorporated.
- Kogut, B., & Zander, U. (1992). Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization Science*, 3(3), 383–397. <http://doi.org/10.1287/orsc.3.3.383>
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META Group Research Note, 6.
- Lavalle, S. (2011). Big Data, Analytics and the Path From Insights to Value. Retrieved from <http://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/>
- Machado, M. P. (2004). A consistent estimator for the binomial distribution in the presence of “incidental parameters”: an application to patent data. *Journal of Econometrics*, 119(1), 73–98. [http://doi.org/10.1016/S0304-4076\(03\)00156-8](http://doi.org/10.1016/S0304-4076(03)00156-8)
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. Retrieved from [http://www.mckinsey.com/Insights/MGI/Research/Technology\\_and\\_Innovation/Big\\_data\\_The\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation)
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & others. (2013). Big data: The next frontier for innovation, competition, and productivity. 2011, McKinsey Global Institute.
- Martínez-Román, J. A., Gamero, J., & Tamayo, J. A. (2011). Analysis of innovation in SMEs using an innovative capability-based non-linear model: A study in the province of Seville (Spain). *Technovation*, 31(9), 459–475. <http://doi.org/10.1016/j.technovation.2011.05.005>
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60–6.
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data. *The Management Revolution*. *Harvard Bus Rev*, 90(10), 61–67.
- Nelson, R. R. (1982). The Role of Knowledge in R&D Efficiency. *The Quarterly Journal of Economics*, 97(3), 453–470. <http://doi.org/10.2307/1885872>
- Nelson, R. R., & Winter, S. G. (2009). *An Evolutionary Theory of Economic Change*. Harvard University Press.
- OECD, & Eurostat. (2005). *Oslo Manual*. OECD Publishing. Retrieved from [http://www.keepeek.com/Digital-Asset-Management/oecd/science-and-technology/oslo-manual\\_9789264013100-en#page1](http://www.keepeek.com/Digital-Asset-Management/oecd/science-and-technology/oslo-manual_9789264013100-en#page1)
- Prior, D. D. (2013). Supplier representative activities and customer perceived value in complex industrial solutions. *Industrial Marketing Management*, 42(8), 1192–1201. <http://doi.org/10.1016/j.indmarman.2013.03.015>
- Quintero, D., Casali, D. de S., Lima, M. C., Szabo, I. G., Olejniczak, M., Mello, T. R. de, ... Redbooks, I. B. M. (2015). IBM Software Defined Infrastructure for Big Data Analytics Workloads. IBM Redbooks.
- Rawlings, N. D., Waller, M., Barrett, A. J., & Bateman, A. (2013). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, gkt953. <http://doi.org/10.1093/nar/gkt953>
- Silveira, M., Marcolin, C. B., & Freitas, H. M. R. (2016). Uso Corporativo do Big Data: Uma Revisão de Literatura. *Revista de Gestão e Projetos - GeP*, 6(3), 44–59. <http://doi.org/10.5585/gep.v6i3.369>
- Taurion, C. (2013). Big Data. Brasport.

Teece, D., & Pisano, G. (1994). The Dynamic Capabilities of Firms: an Introduction. *Industrial and Corporate Change*, 3(3), 537–556. <http://doi.org/10.1093/icc/3.3.537-a>

Tian, L., Zhiping, Y., & Zhengyin, H. (2013). The Large Aperture Optical Elements patent search system based on Domain Knowledge Organization System. *World Patent Information*, 35(3), 209–213. <http://doi.org/10.1016/j.wpi.2013.04.007>

Tien, J. M. (2013). Big Data: Unleashing information. *Journal of Systems Science and Systems Engineering*, 22(2), 127–151. <http://doi.org/10.1007/s11518-013-5219-4>

Vasarhelyi, M. A., & Halper, F. B. (1991). The continuous audit of online systems. *Auditing: A Journal of Practice & Theory*, 10(1), 110–125.

Volpato, T., Rufino, R. R., & Dias, J. W. (2014). BIG DATA—TRANSFORMANDO DADOS EM DECISÕES. Retrieved from

[http://web.unipar.br/~seinpar/2014/artigos/graduacao/Tiago\\_Volpato.pdf](http://web.unipar.br/~seinpar/2014/artigos/graduacao/Tiago_Volpato.pdf)

Waller, M. A., & Fawcett, S. E. (2013). Click Here for a Data Scientist: Big Data, Predictive Analytics, and Theory Development in the Era of a Maker Movement Supply Chain. *Journal of Business Logistics*, 34(4), 249–252.

Williamson, O. E. (1981). The Economics of Organization: The Transaction Cost Approach. *The American Journal of Sociology*, 87, 548–577.

Xindong Wu, Xingquan Zhu, Gong-Qing Wu, & Wei Ding. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97–107. <http://doi.org/10.1109/TKDE.2013.109>

Yugue, R. T., & Maximiano, A. C. A. (2013). Understanding and Managing Project Complexity. *Revista de Gestão E Projetos - eISSN: 2236-0972*, 4(1), 01–22. <http://doi.org/10.5585/gep.v4i1.109>

### Cite it like this:

Mazzieri, M., & Soares, E. D. (2016). Conceptualization and theorization of the Big Data. *International Journal of Innovation (IJI Journal)*, 4(2). doi:<http://dx.doi.org/10.5585/iji.v4i2.91>