BIG DATA ANALYTICS: ACHIEVEMENTS, CHALLENGES, AND RESEARCH TRENDS

Andre Coelho Vaz Henriques
Fundacao Getulio Vargas, Brazil
E-mail: acvhenriques@gmail.com

Fernando de Souza Meirelles
Fundacao Getulio Vargas, Brazil
E-mail: fernando.meirelles@fgv.br

Maria Alexandra Viegas Cortez da Cunha
Fundacao Getulio Vargas, Brazil
E-mail: alexandra.cunha@fgv.br

ABSTRACT

Big data applications combined with analytical tools foster prediction techniques that impact societal, economic, and political changes. After almost a decade of studies, this paper proposes to identify major debates on big data analytics, presenting its evolution over the past years and identifying its research tendencies. We limited our research to the top eight journals in information systems. Our findings suggest that big data analytics is apparently reaching a plateau, which might be confirmed by publications in the following years. The paper contributes to the current debate on big data by identifying ongoing studies in the research community. In addition, it provides a critical analysis of the field development, from its perceived benefits to its unimagined consequences. Finally, we conclude that other perspectives on big data analytics might include a new wave of studies and that new paths beyond productivity gains can be explored.

Keywords: Big data. Analytics. Business Intelligence. Datification. Data Science.

1. INTRODUCTION

The term big data refers to data whose size goes beyond the ability of regular database software to capture, store, manage, and analyze (MANYIKA et al., 2011). Big data applications combined with analytical tools (or big data analytics) foster prediction techniques that influence societal, economic, and political changes. After almost a decade of studies, this paper proposes to identify major debates regarding big data, presenting big data's evolution over the past years and identifying its research tendencies.

In the big data era people are computer-mediated in their daily activities, generating a very large amount of digital records. Additionally, in a socioeconomic environment heavily influenced by mobile applications (apps), each transaction involving any text, digital procedure, tactile command, voice, and other user inputs in an app is data. This context presents a myriad of possibilities to take advantage of big data analytics. The success of companies such as Google, eBay, Facebook, and Amazon arouses interest and draws attention to the big data phenomenon both in the academic and business worlds. These corporations, just to name a few, are the hallmarks of big data applications.

But the advances originating from BDA technologies raise new issues. On the political side, the 2016 elections in the United States were strongly affected by media resources based on BDA—the same techniques were employed more recently in Brazil, exerting a major influence on the results of the 2018 elections. According to The Economist (2018), the Chinese government is working on a surveillance system based on facial recognition, including factors such as emotions and sexuality, aiming to control its population in an unprecedented way. On the other hand, San Francisco, in California—the center of the technology revolution—just banned facial recognition by police and certain agencies (THE NEW YORK TIMES, 2019). These types of developments evoke questions, such as the limits of privacy over other interests, which need further discussion.

While the strategic value of data processed by algorithms promotes great efficiency to corporations, the implications for society and individuals are not clear. Decision models leveraged by sophisticated algorithms can replace the judgements of complex analyses, invading knowledge occupation professions. Therefore, jobs, institutions, and industries established today might be affected in uncertain ways. These enabling technologies may modify markets all over the world, leading to impacts that are still unknown. Thus, the technology's many benefits can lead to negative consequences.

1202

In the current scenario, a systematic analysis of the field's evolution, clarifying topics that have already been investigated and pointing out issues that still need further research, is lacking. To fill that gap, the following research question is asked: what are the current debates in big data analytics field and what are its research trends? The study synthesizes major challenges and concerns regarding BDA, presents the field's development over time, and points out gaps that need further investigation. Although research has been conducted in this area, the present analysis, based on the eight major journals on information systems (IS), provides a new perspective.

From an academic point of view, this study presents a clear picture of BDA development over time and uncovers gaps that have not yet been addressed. Such analysis provides a better understanding of big data analytics applications and their consequences, generating reflections on the possibilities and boundaries in the field. For practitioners, the study brings together BDA techniques, models, and a mindset that have been successfully applied, and at the same time, it provides a warning regarding BDA limits and brings attention to its applications.

To achieve this, we first present a theoretical foundation, introducing the concepts of big data analytics. Then, we describe the method applied in the research. Next, we discuss the production of articles, the impact and challenges generated by BDA, a retrospective of major contributions, and expectations of new studies on the topic. Finally, we present our conclusions, including the limitations of the study and future research suggestions.

2. EXISTING THEORETICAL FOUNDATIONS

The propagation of the web, social media, mobile apps, and sensor networks, in addition to the cost reduction in storage and computing resources, has given rise to ubiquitous and increasing digital computer records termed big data (MULLER et al., 2016), while the use of analytics to extract value from big data has given rise to big data analytics (MULLER; FAY; BROCKE, 2018).

Computers embedded in products such as cars, vacuum cleaners, or video consoles have given rise to large amounts of digitized data (LOEBBECKE; PICOT, 2015). Location-based processes and the internet of things also contribute to data generation (LYYTINEN; GROVER, 2017); therefore, with all these resources, technology provides the opportunity to transform data into 'actionable insights' (SABOO; KUMAR; PARK, 2016; KITCHENS et al., 2018).

1203

That is, BDA emerged to describe the analytical technologies employed and the large and complex amounts of data required to manage them.

The term intelligence has been used in academic literature since the 1950s, but only in the 1990s has it become popular in business and IT communities (CHEN; CHIANG; STOREY, 2012). Hoping not to commit a heresy, we understand the big data analytics concept as being very similar to the more famous (and 'less sexy')—as put by Newell and Marabelli (2015)—business intelligence (BI).

Hence, one can consider big data as a close successor of business intelligence (ABBASI; SARKER; CHIANG, 2016). In other words, the term 'big data analytics' (or just big data) has been adopted to refer to data sets and analytical techniques for large and advanced applications, requiring complex techniques in their usage.

We are going through a transition period in which limited volume, regular velocity, and small variety are being replaced with a new concept of information that is very different from the traditional one, as precisely described by Abbasi, Sarker and Chiang (2016).

In this context, the central role is played by structured data, which is stored in data centers employing relational database management systems (RDBMS). In this arrangement, many organizations integrate structured data sources in data warehouses and data marts that use extract, transform, and load (ETL) technologies. The stored data are analyzed by data analysts and programmers using structured query language (SQL), reverting the resulting data to BI tools, report generators, or analytical models employed in predictive technologies.

Conversely, in the knowledge stage of the value chain, there are direct interactions among the information enablers and decision makers, that is, the consumers and producers of information. In this scenario, technologies such as knowledge management systems, corporate wikis, BI dashboards, reporting tools, and expert systems revert knowledge through technologies such as decision support systems (DSS) collaboration tools and recommender systems that guide decision-making processes by analysts and managers (ABBASI; SARKER; CHIANG, 2016).

Thus, BDA has significatively advanced from its early stage of business intelligence 1.0—marked by structured data, dashboards, data mining, OLAP, and statistical analyses; to 2.0—distinguished by unstructured online data, social network analyses, web analytics and intelligence, and social media analytics; until the current 3.0 era—characterized by mobile and

1204

sensor-based content, mobile analytics and location, and context relevant analyses (CHEN; CHIANG; STOREY, 2012; GROVER et al., 2018).

Therefore, BDA today not only involves well-structured traditional data stored in traditional databases and data warehouses (BAESENS et al., 2016) but also implies large, diverse, and dynamic sets of digital traces and user-generated content in addition to analytics methods (MULLER et al., 2016). It encompasses public, proprietary, and purchased sources of unstructured data, including documents, web content, video, image, audio, and sensor data (GROVER et al., 2018) whose development is far from trivial (CONSTANTIOU; KALLINIKOS, 2015).

It involves the analyses and interpretations of all kinds of digital information (LOEBBECKE; PICOT, 2015) and arises from major sources, including large-scale enterprise systems, online social graphs, mobile devices, the internet of things, and open data (BAESENS et al., 2016). It borrows techniques grounded in statistics, machine learning, and econometrics, among others.

There are some big data features and functionalities that are commonly called management 'V's: volume, variety, and velocity (CHEN; PRESTON; SWINK, 2015; MULLER et al., 2016; CLARKE, 2016; ABBASI; SARKER; CHIANG, 2016; HAN; PARK; OH, 2016), which means data that are too large, fast, or hard to process. Volume refers to the enormous amount of data to be processed.

Velocity refers to the necessity of the speed with which data are processed, from their generation to their use. One of the most challenging aspects of this chain maybe be the time from data extraction until the generation of value from the data, that is, when the data becomes useful or relevant (CONSTANTIOU; KALLINIKOS, 2015; BAESENS et al. 2016).

Variety is related to the great diversity of origins, forms, and formats of data, which makes them difficult to categorize and tabulate. It involves not only traditional data but also user-generated text, videos, images, social network data, web and mobile clickstreams, sensor-based data, and spatial-temporal data (MCAFEE; BRYNJOLFSSON, 2012; ABBASI; SARKER; CHIANG, 2016).

More recently, some authors have incorporated other 'V's in this hall, such as 'variability' and 'value'. Variability is related to the susceptibility of the data to changes, such as when it is translated into another language (NUAMI et al., 2015). In terms of value, the concept of big data involves not only a vast amount of data but also the process by which

organizations derive value from them—which inevitably varies across organizations, situations, and managers (LYYTINEN; GROVER, 2017), e.g., improving organizational decision making, promoting service innovation, and ensuring higher satisfaction and retention (GROVER et al., 2018). Last, Abbasi, Sarker and Chiang (2016) and others consider another 'V' in the information value chain—veracity, which refers to the truthfulness of the data.

In BDA, advanced technologies are employed to analyze data to discover useful information that is hidden, such as unknown correlations, or to uncover patterns (CHEN; PRESTON; SWINK, 2015), providing answers to questions that have not even been considered (GROVER et al., 2018). In contrast to research in which data are collected for a specific end and measured by validated instruments, big data often just happens (MULLER et al., 2016).

Since large samples are becoming more common in the IS field, researchers are increasingly working with big data (CHATLA; SHMUELI, 2017). However, Zuboff (2015) criticizes the passive position assumed regarding the topic, saying that the literature's view of BDA as a technological phenomenon disregards its social origin. On Zuboff's view, big data have an intentional sense and severe consequences, predicting and modifying human behavior through a logic that he refers to as 'surveillance capitalism'.

In fact, aspects such as privacy, surveillance, and democracy arouse debates that still need further investigation. In this sense, digitization and big data analytics, or 'datification' (GALLIERS et al., 2015; NEWELL; MARABELLI, 2015; LOEBBECKE; PICOT, 2015), are embedded in all areas of life. Interactions with objects with sensors and IP addresses provide a mass of data sources, and humans have become 'walking data generators' (MCAFEE; BRYNJOLFSSON, 2012; LOEBBECKE; PICOT, 2015).

3. RESEARCH APPROACH

To address the aim of this paper, we carried out an analysis of the field's evolution, clarifying topics that have already been investigated and pointing out issues that still need further research. The literature review adopted was concept-centric. According to Webster and Watson (2002), in this kind of review the concepts that determine the organizing framework in order to synthase the literature.

To provide a systematic review, we limited our research to the top eight Journals based on the Association for Information Systems (AIS) Senior Academic Collegiate—the most respected and recognized Journals in Information Systems field. It includes the European Journal of Information Systems, the Information Systems Journal, Information Systems

1206

Research, the Journal of AIS, the Journal of Information Technology, the Journal of MIS, the Journal of Strategic Information Systems, and MIS Quarterly. Instead of a longitudinal analyses of a vast number of papers, the aim was to deeply analyze the articles, which contemplate geographical, methodological, and topic diversity considerations.

The list includes the mature and established knowledge, being representative of the IS field. Thus, the 'Basket of Eight' Journals can reflect the core body of knowledge in IS, serving as a data source for investigating the field's development.

We assumed that not all the papers discussing 'big data' adopt that specific term. Therefore, we first expanded our search by looking for papers containing the terms 'analytics' and 'intelligence' (CHEN; CHIANG; STOREY, 2012; LUVIZAN; DINIZ, 2017) in the keywords, title, or abstract. However, the papers analyses showed that 'datification' and 'data science' are quite common in related fields, which led us to include both these terms in our search.

In an initial analysis, the papers were selected, discarded, or subjected to a fine-grained examination. Articles explicitly containing the term 'big data' in any of the search fields were included in the study, while all the others went through a verification process. The articles that did not adopt specific 'big data' nomenclature but that combined large, diverse, and dynamic data—with broad academic consensus determining the conditions under which big data emerges—were included. As technology progresses over time, the size of the dataset classified as big data increases.

For this reason, authors like Manyika et al. (2011) understand that it would not be reasonable to define a quantity of data that characterize it. In line with these authors, the criteria adopted to determine a large volume of data was those unable to be managed by regular software, such as Excel and the like. Articles discussing regular databases, enterprise systems (e.g., ERP, CRM, e-commerce) and traditional predictive analytics, among others, were discarded, and those that somehow contributed to clarifying big data aspects (e.g., advanced text analytics tools) were included. All the selected articles were subjected to a rigorous reading and were classified in a spreadsheet.

The time limit defined was the year 2010, since we understand that the big data phenomenon only emerged after enabling technologies arose from that year onwards. The papers were collected between November 1st of 2018 and January 15th of 2019, and the searches contemplated the articles published between 2010 and 2018.

1207

However, it was not found articles published in 2010 and 2011. Out of the 135 candidates in the initial pool, we selected 41 papers that met the selection criteria. Our analyses focused on summarizing the main findings of the papers, highlighting current debates, and finding aspects that could characterize and classify the articles. Nevertheless, the intention of this study was not to merely describe the area but to actually contribute to new research, pointing out gaps and trends in the literature.

4. DISCUSSION

4.1. Global Production

The first factor that determined the selection of articles in this research was the country with which the authors were associated at the time of publication. As we can see in the next figure, the publications come almost entirely from the northern hemisphere; Australia is the only exception. We see no publications on BDA at all from South America or the entire African continent, and we find that authors whose institutions are based in the United States (26 of 60) produced nearly half the publications.

Figure 1: 'Basket of Eight' BDA World Production
Source: created by the authors

After the United States, the countries of China, the United Kingdom, and the Netherlands are tied with four publications, followed by Denmark, Liechtenstein, and Taiwan, with three publications. Germany, Hong Kong, India, Israel, and South Korea follow with two, and finally, there are Australia, Belgium, and Switzerland with one publication each. We assume that this scenario—in which publications originating from the United States and European countries prevail—is probably not unique to BDA publications but reflects the continuum of global production in the 'Basket of Eight'.

1208

One fact, however, attracts our attention. Except for China and India, the other members of BRICS (Brazil, Russia, and South Africa) also have no publications in the field in the leading journals. This fact catches our attention, considering the size and economic influence of the countries that compose the BRICS. One major potential of BDA is precisely the fostering of economic gains—not to mention all the social and political aspects.

Conversely, countries with a more modest global presence—such as Liechtenstein, Taiwan, and Israel—share the stage with large, developed nations. The notable accomplishments of these countries perhaps encourage professionals and academics from the rest of the world which have not yet reached such a 'title' in the field.

Another interesting fact observed regards the institutions with which the authors were associated when publishing their articles related to BDA. The only institution that published four times in the leading journals in the field was the University of Liechtenstein, from the Principality of Liechtenstein.

The monarchy is situated between Austria and Switzerland and has a population of nearly 38,000, and its University is the leading producer of BDA in the 'Basket of Eight'. The intention behind this global overview of big data publications in leading journals is to provide a big picture of the field and of how efforts could be directed or rethought.

4.2.   Impacts and Challenges Generated by Big Data Analytics

Data quality, analytical tools, and human analytics talent are some of the enablers of BDA that generate insights and valuable knowledge for decision making. Moreover, while BDA opens new opportunities, it also introduces new challenges, such as the impact on the labor market and privacy concerns. Regarding this topic, we analyze the major impacts and challenges identified in the reading of the papers from the 'Basket of Eight' journals regarding big data analytics.

To provide a big picture of the main issues faced by the field, we first present the following figure with keywords extracted from the selected articles in this research. The keywords are ranked according to their frequency and displayed in a tag cloud visualization. The terms 'big data', 'big data analytics', 'analytics', and 'business intelligence' were removed to highlight topics published in the articles selected in this research.

Figure 2: Keywords Tag Cloud
Source: created by the authors

The diversity may reflect the ramifications of the topic, from which different paths emerge. Nevertheless, some words also attract attention, such as data quality, privacy, social data, and sentiment analyses.

In the following, we highlight some challenges and concerns in the field, based on the common topics referred to in the selected articles.

4.2.1. Qualified Professionals

The benefits promoted by big data applications are diverse. However, in addition to technology, BDA adoption requires qualified professionals. Both academic and nonacademic literature points out that the shortage of professionals and individuals capable of using the big data potential may be one of the major difficulties in its application and development. Qualified professionals with experience and expertise are key to developing and implementing BDA strategies, including data scientists, programmers, developers, and analysts (GROVER et al., 2018).

According to Baesens et al. (2016), most universities do not offer mature programs and classes on BDA, and even worse, many professors do not have the necessary knowledge to effectively deliver big data education. To solve this issue, the authors conclude that alliances between the academy and the business world could help provide good quality education programs.

4.2.2. Privacy

Although the advantages of the network economy are well known, concerns about privacy have emerged in research. Big data analytics technologies move faster than the chain of systems that preserves privacy and information security (LOWRY; DINEV; WILLISON, 2017).

1210

Zuboff (2015) points out that today, we have data from several sensors embedded in objects, bodies, and places. The author draws attention to the fact that some technology companies put innovation first and disregard the consequences—e.g., exhibiting a photograph of a private property without license.

Apparently, users have been persuaded to ignore the dark side of datification and its package of digital traces because the benefits are higher than the costs. Therefore, it seems that individuals perceive that it is better to be able to look for something specific on Google (and thus support the algorithm that knows about what we want and about us) than to simply not use it (NEWELL; MARABELLI, 2015).

Aligned with this, the General Data Protection Regulation (GDPR) enacted in 2016 is a European law that enhances data protection for Europe's citizens and thus ensures that all small, medium, and large companies will have to invest in cybersecurity. In addition to local companies, companies all over the world that have business with Europe need to adjust to this regulation. Similarly, the 'right to be forgotten' was sanctioned by the European Union court in 2014—that is, links to 'irrelevant' or 'outdated' information may be deleted whenever requested by citizens of the European Union.

4.2.3. Little Data

While big data are data originating from indiscriminate groups with the logic of decision-making algorithms, a recent phenomenon called 'little data' might be emerging (NEWELL; MARABELLI, 2015). It uses big data to direct knowledge in a targeted way that is potentially unfair, predicting the behavior of a particular individual. These kinds of actions might have a serious impact, giving rise to questions about the boundaries of BDA in ethical and privacy domains. Digitized devices that are able to track and record individuals' actions permeate our lives and pose relevant questions that still need to be addressed.

4.2.4. The Labor Market

The replacement of humans by machines in basic and routine activities is not a recent phenomenon. Now, machines are progressively starting to replace humans in cognitive tasks, since big data-based systems are becoming more cost effective and have a higher hit rate (LOEBBECKE; PICOT, 2015). The consequences of this change are still obscure, but it seems that it will dramatically modify the current configuration of several professions.

1211

4.2.5. Algorithm Complexity

Although sometimes algorithms are very good at predictions, in several cases they are incomprehensible (MULLER et al., 2016). It is necessary to understand the relation between data and the analyzed phenomenon (LYYTINEN; GROVER, 2017). Nevertheless, highly advanced algorithms composed of complex formulas closed in black boxes are unlikely to be adopted to support key strategic business areas, such as fraud detection, credit risk measurement, or medical diagnosis (BAESENS et al., 2016).

4.2.6. Infrastructure

Big data analytics infrastructure implies the collecting of different types of data, sharing data, and integrating sources of data. In addition to human talent, organizations need to invest in analytics portfolios and big data assets to promote their development.

BDA infrastructure encompasses data sources (e.g., clickstream, transactional, user-generated, social media) and proper platforms to collect, ingrate, share, process, and manage big data—especially those dealing with unstructured data in multiple formats (GROVER et al., 2018). More recently, data lakes have become the current best practice solution to data collection and integration (KITCHENS et al., 2018).

They consist of vast repositories in which organizations store data in their native format until they analyze and extract value from it. This solution reduces costs for sharing data within a firm and promotes experimentation and discoveries. In addition, due to the large size of data, increasingly more outsourced firms work as servers in the so-called 'cloud' (LOWRY; DINEV; WILLISON, 2017).

4.2.7. Data Quality

Park et al. (2012) highlight the fact that little attention has been paid to problems regarding erroneous data by academic research, although institutional agencies such as the US Census Bureau are making efforts in this area. Without proper data quality, resources will inevitably be misallocated (CLARKE, 2016). More often than not, data are noisy, erroneous, and missing; due to exponential growth, ensuring trustworthy sources of data and information is difficult (GROVER et al., 2018).

4.3. Retrospective of Major Contributions

To the best of our knowledge, the publication of the remarkable article by Chen, Chiang and Storey (2012) is a hallmark of big data in IS, clarifying concepts, channeling the term, and

1212

providing guidance for future studies. This paper identifies the evolution, applications, and emerging research areas of BI&A (1.0, 2.0 and 3.0). In the same year, Chau and Xu (2012) developed a technique to effectively collect, extract, and analyze blogs related to a specific topic, and Park et al. (2012) created an inference model based on patterns of social ties that assess the validity of self-reported customer profiles.

In the following years, the big data analytics potential was explored in business. This exploration started to show the implications for strategy making (CONSTANTIOU; KALLINIKOS, 2015) and demonstrated that its adoption influences business growth (CHEN; PRESTON; SWINK, 2015).

Moreover, Constantiou and Kallinikos (2015) focused attention on unstructured data—such as the media of text, image, and sound—which cross the alphanumeric systems that have prevailed in organization management. Additionally, in 2015, the first studies pointing out big data analytics consequences were published, and the terms 'datification' and 'digitization' emerged.

In this sense, Loebbecke and Picot (2015) demonstrate the side effects of big data analytics in business and society; Newell and Marabelli (2015) show the economic, legal, organizational, ethical, cultural, and psychological consequences of digitization—including issues related to privacy, control and dependence; and Zuboff (2015) questions the new global architecture of computer mediation.

However, the year of the big data was 2016. Almost 40% of the articles in the 'Basket of Eight' were published in 2016. Considering the growth of publications and interest in the topic, several studies guiding BDA research gained space.

In an editorial in the Journal of the Association of Information Systems, Abbasi, Sarker and Chiang (2016) discuss the emerging implications for theory and methodology arising from big data's disruptive effects. In line with this, in that very year, MIS Quarterly published its second editorial related to big data (RAI, 2016)—drafting opportunities for IS research—and published a special issue on BDA, leveraging the number of articles on the topic.

Moreover, Ketter et al. (2016) present a conceptual and methodological approach by which IS research can address BDA issues, while Baesens et al. (2016) provide a perspective on emerging research opportunities regarding big data, and Muller et al. (2016) set guidelines for conducting BDA studies in IS.

1213

At the same time, 2016 is also marked by studies introducing new models and techniques. In this regard, we highlight the works of Brynjolfsson, Geva and Reichman (2016), who demonstrate a crowd-squared approach for predicting search trend data; Lash and Zhao (2016), who create a system able to predict movie profitability in the preproduction stage; and Shi, Lee and Whinston (2016), whose works enhance decision making in mergers and acquisitions through BDA techniques.

Furthermore, Menon and Sarkar (2016) present a scalable approach to solve privacy concerns when sharing transactional databases, and Li, Chen and Nunamaker (2016) develop a system that is capable of identifying underground economy sellers. Finally, Clarke (2016) draws attention to the moral and legal responsibilities of computing researchers and professionals.

Apparently, big data analytics publications reached their peak in 2017. The large number of publications was replaced by a reduced (but not less notable) quantity of articles. In fact, the works produced in 2017 brought novel insights. Kelly and Noonan (2017), with the Indian public health service, show how systematic practices of working with data prevail and how the challenge of conceiving new forms of data continues to appear in familiar ways.

Furthermore, Guo et al. (2017) innovate with a system framework capable of extracting a small number of articles that could represent the diversified content generated on an organizational blogging platform. Finally, Gunther (2017) clarifies how organizations realize value from big data—a concept further investigated by Müller, Fay and Brocke (2018)—providing objective estimations of BDA business value.

The publications of 2018 were marked by a few exotic studies and novel contributions. In this regard, we mention the work of Aversa, Cabantous and Haefliger (2018), in which, by means of a Formula 1 race, the authors determine that decision support system (DSS) potential failure is exacerbated under pressure and time constraints.

Additionally, Deng et al. (2018) and Li, Dalen and Rees (2018) analyze sentiment within big data. The former authors show the influence of microblog sentiment on stock returns, while the latter verify that stock microblog features serve as proxies for market sentiment. Furthermore, Lehrer et al. (2018) clarify how BDA technologies enable service innovation, and Zhou et al. (2018) identify the limits of BDA; they verify that increasing review volume reduces customer agility.

1214

Based on the number of publications and exploration of the field, it seems that BDA is reaching a plateau—which might indicate its maturity level. The next figure shows the number of publications per year, demonstrating the evolution of the field in terms of articles published in the 'Basket of Eight'.
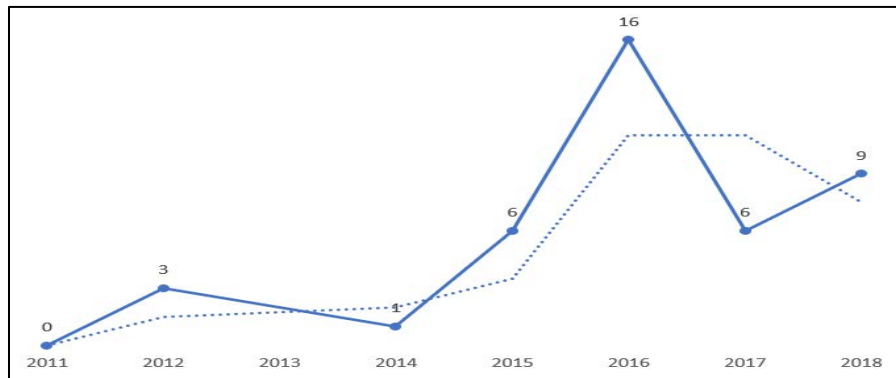

Figure 3: BDA Production in the 'Basket of Eight'
Source: created by the authors

Given the context presented, we opted to compare this evidence within an expanded scenario. Searching for the words 'Big Data Analytics' on Google Trends we found signs that corroborates to the possible plateau the technology might be reaching. As shown in the next figure, based on the production between 2010 and 2018, an inflection point takes place in 2017 and a tendency of a decreasing number of articles in the field—which might be confirmed in the following years.
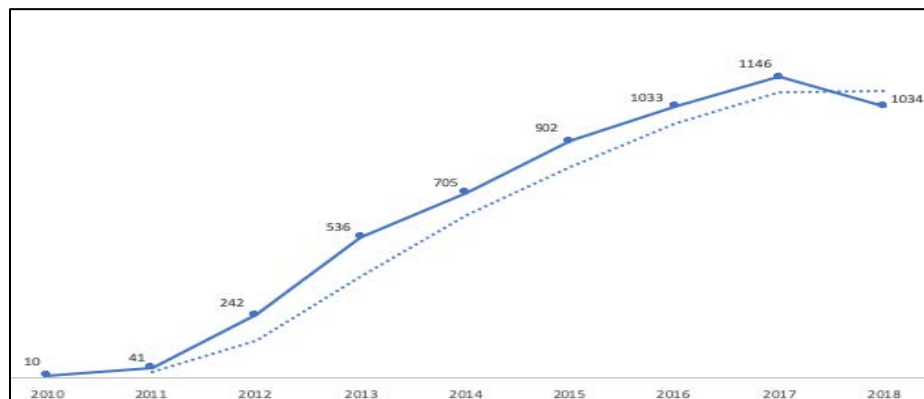

Figure 4: BDA Production according to Google Trends
Source: created by the authors—based on Google Trends data

The fact that it might be reaching a plateau does not mean, however, that the field is fully explored. Instead, it might show the maturity of the technology.

Based on the published content since 2010 in the 'Basket of Eight', it is possible to identify different waves of BDA, as shown in the following figure. The analyses clarify diverse events that include BDA's first studies, potential for business, social media data, consequences,

research concerns, information security and privacy concerns, new models and techniques, sentiment analyses, and finally (what appears to be) a plateau.
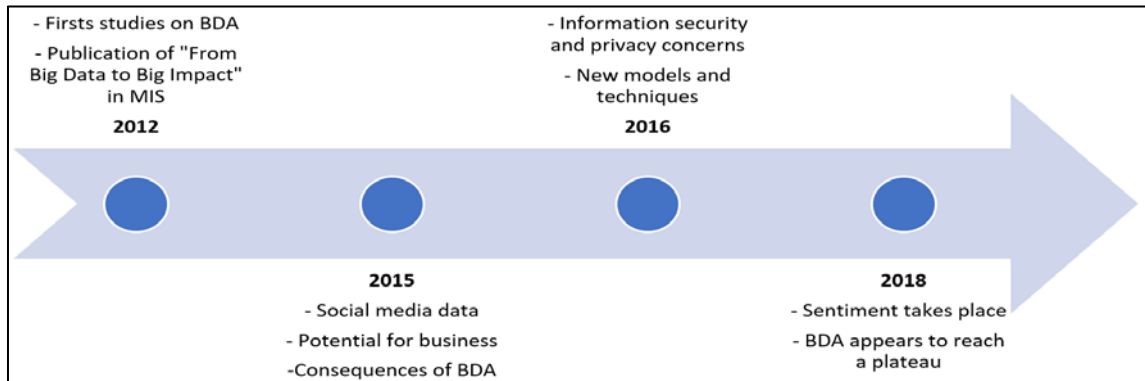

Figure 5: BDA Evolution
Source: created by the authors

4.4.   Research Trends

Several highlights from articles from a few years ago have already been addressed, which leads us to focus on those ideas we consider more relevant and that are still in need of further research. We also choose not to highlight issues regarding specific topics from other areas (e.g., mergers and acquisitions, the stock market, customer behavior); without denying the value of these formidable works, their scope goes beyond the IS field.

Therefore, we try to indicate future studies in a broader sense, bringing out findings that may be applicable in the information systems field as a whole. Similarly, we do not focus on broader variations of similar studies (e.g., allowing the generalizability of research results, enhancing study validity, or approaching other—but similar—dimensions or domains). Rather, we mostly choose insights that we believe somehow shake up BDA in the IS field. In the following table, we compile promising research opportunities on this topic based on the analysis of the selected articles.

Table 1: Research Opportunities

Research Opportunities	Brief Description	Authors
Theories and Methods	BDA is not merely a data process change but is highly disruptive for academic studies, making it necessary to reassess our research methodologies, assumptions, and substantive questions.	ABBASI; SARKER; CHIANG (2016); BAESENS et al. (2016); LYYTINEN; GROVER (2017).
Interdisciplinary Studies	Researchers should consider collaborating with other areas, which could result in advancing the IS field through the introduction of new methodological tools.	AVERSA; CABANTOUS; HAEFLIGER (2018); BREUKER et al. (2018); GUNTHER et al. (2017); LOEBBECKE; PICOT (2015); MULLER et al. (2016).

1216

Privacy, Ethics, Security, and Surveillance	There is a need for studies on surveillance by private and public authorities, which includes the protection of individual rights, privacy, ethical issues, and risk concerns.	BREUKER et al. (2018); GUNTHER et al. (2017); LOWRY; DINEV; WILLISON (2017); ZUBOFF (2015).
Service Innovation	There are missing studies on approaching BDA materiality and how it enables service innovation.	KELLY; NOONAN (2017); LEHRER et al. (2018).
New BDA Applications	There are several research opportunities regarding BDA applications, including sentiment, perspectives from outside the data, and the meaning and relevance of images and videos.	AVERSA; CABANTOUS; HAEFLIGER (2018); CONSTANTIOU; KALLINIKOS (2015); DENG et al. (2018); GUO et al. (2017); SABOO, KUMAR; PARK (2016); KITCHENS et al. (2018).
Governance	There is a need to broaden our understanding of information governance, identifying how antecedents (enablers or inhibitors) apply to it and how it affects organizational performance.	TALLON; RAMIREZ; SHORT (2014)
Social Impacts	Studies on the broad social issues raised by BDA are missing, including how digitization (as an actor) affects social relationships.	LOEBBECKE; PICOT (2015); NEWELL; MARABELLI (2015).
BDA Value	There is still a gap in reliable empirical evidence on BDA's business value, making it necessary to explore how organizations effectively convert big data potential into economic and social value.	ABBASI; SARKER; CHIANG (2016); GROVER et al. (2018); GUNTHER et al. (2017); MULLER; FAY; BROCKE (2018).

Source: created by the authors

5. CONCLUSION

After a peak in publications in 2016, it appears that BDA will soon reach a plateau—which might be confirmed by publications in the following years. In part, it may be that BDA is being replaced by new terminologies (e.g., data science, datification) but mostly it is being transformed to have new, complex and deep ramifications.

It seems that we are arriving at a land of big data impacts. We are going through a transition in which a new analytical mindset is taking created place, and the boundaries of what we can and cannot do are still obscure.

Given the availability of data, different kinds of devices, machine learning, algorithms, sensors, and data clouds provide endless possibilities. Many solutions have been found. Perhaps other perspectives can now be more deeply explored. According to the papers analyzed in this research, there are topics that still need further investigation. In this regard, we highlight privacy concerns and ethical aspects, impacts on society, new applications, and interdisciplinary research that might constitute new waves of studies, defining and limiting BDA boundaries.

Concerning the privacy and ethical aspects, no one wants to live in a 'Big Brother' environment, but we all want the privileges that the 'sharing' of data allows. There is a need to

1217

revisit various aspects of the social pact with technology, considering where more transparency and information are needed. What is our relationship to data and how can it help or harm us? People need to understand where they are heading and what big data means for the market. When accepting cookies to access certain data, for instance, how many people actually know what a cookie is? There is a need to educate and inform people to make them understand the tradeoffs that come with the data that they provide.

Furthermore, most of the related works focus on increasing efficiency, mainly on supporting the private sector. Perhaps opportunities to explore the benefits for society and other areas are being left behind. How can BDA effectively help people's lives in cities? How might BDA help with water consumption in less developed regions, agriculture, or governments—how can it generate value for society? In part, these results may have been found because of the nature and purpose of the searched journals.

However, the above are still issues that might be more deeply explored. That is, studies on BDA could explore how to improve people's quality of life, not just how to increase business results. We mean that big data analytics can go beyond cost reduction, optimization, productivity gains, increased efficiency, and so on—by providing analyses from a social perspective.

In addition, it seems that new techniques will form a continuum in BDA, especially in congregating data. We understand that integrating silos of data might be a fruitful path to explore. Future works might expand the area through collaborations of IS academics and professionals in other fields, integrating advances such as machine learning and human interaction and developing systems to integrate others.

In addition, the absolute absence of publications from South America and Africa, as well as the modest participation of BRICS, whereby Brazil, Russia, and South Africa are still mute in the leading journals, is frightful. Professionals, researchers, and even government agents from these large nations might lose the opportunity to explore a field full of possibilities. We hope that this finding encourages them to expand their research in this area. At the same time, the University of Liechtenstein, for example, might be an outstanding place for the development of data science professionals.

This study contributes to the academy by synthesizing major challenges and concerns regarding big data analytics, presenting its evolutionary waves and development over time and indicating research tendencies that can be further explored—and that go beyond business

1218

efficiency. For practitioners, it presents techniques and models that have been successfully applied and that are rapidly being disseminated. At the same time, it warns about the limits of BDA and draws attention to issues that should be considered.

The more that technologies develop, the more possibilities there are. This might be an endless race: each time faster, each time better. Big data analytics are good for those who produce and for those who consume. However, this does not give us the right to ignore the impact that the technology generates.

Debates regarding machines taking our jobs are pertinent and essential, of course, but this is yet another chapter of the industrial revolution—which is now taking place by means of other kinds of technology. Further debates and studies are needed to understand (and forecast) changes and to define proper boundaries—whether through ethical, cultural, legal, or other means. When the elevator was invented, the obligatory position of the elevator operator was created. Disruptive technologies go through this process of acceptance in various spheres of society.

Last, although this research accomplishes the aim of providing a broad picture of BDA in the most acknowledged journals, this study is limited by the method adopted, as it analyzes only the eight major journals in IS. More studies expanding this perspective could provide a broader view of the field. Additionally, future studies could adopt other methods of content analysis to treat the data collected, such as semantic, morphological, structural, syntax, among others.

REFERENCES

ABBASI, A.; SARKER, S.; CHIANG, R. H. L. (2016) Big Data Research in Information Systems: Toward an Inclusive Research Agenda. **Journal of the Association of Information Systems,** v. 17, n. 2, p. 1–32.

AVERSA, P.; CABANTOUS, L.; HAEFLIGER, S. (2018) When Decision Support Systems Fail: Insights for Strategic Information Systems from Formula 1. **Journal of Strategic Information Systems,** v. 27, n. 3, p. 221–236.

BAESENS, B.; BAPNA R.; MARSDEN J. R.; VANTHIENEN, J.; ZHAO J. L. (2016) Transformational Issues of Big Data and Analytics in Networked Business. **MIS Quarterly,** v. 40, n. 4, p. 807-818.

BREUKER, D.; MATZNER, M; DELFMANN, P.; BECKER, J. (2016) Comprehensible Predictive Models for Business Process. **MIS Quarterly,** v. 40, n. 4, p. 1009-1034.

BRYNJOLFSSON, E.; GEVA, T.; REICHMAN, S. (2016) Crowd-Squared: Amplifying the Predictive Power of Search Trend Data. **MIS Quarterly,** v. 40, n. 4, p. 941-961.

CHATLA, S. B.; SHMUELI, G. (2017) An Extensive Examination of Regression Models with a Binary Outcome Variable. **Journal of the Association for Information Systems,** v. 18, n. 4 p. 340–371.

CHAU, M.; XU, J. (2012). Business Intelligence in Blogs: Understanding Consumer Interactions and Communities. **MIS Quarterly,** v. 36, n. 4, p. 1189-1216.

CHEN, M.; WANG, P. (2018) A Roadmap to Determine the Important Factors of the House Value: A case study by using actual price registration data of Taipei housing transactions. **Independent Journal of Management and Production,** v. 9, n. 1, p. 245-261.

CHEN, D. Q.; PRESTON, D. S.; SWINK, M. (2015) How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management. **Journal of Management Information Systems,** v. 32, n. 4, p. 4–39.

CHEN, H.; CHIANG; R. H.; STOREY, V. C. (2012) Business Intelligence and Analytics: From Big Data to Big Impact. **Journal of Management Information Systems Quarterly,** v. 36, n. 4, p. 1165-1188.

CLARKE, R. (2016) Big Data, Big Risks. **Information Systems Journal,** v. 26, n. 1, p. 77–90.

CONSTANTIOU, I. D.; KALLINIKOS, J. (2015) New Games, New Rules: Big Data and the Changing Context of Strategy. **Journal of Information Technology,** v. 30, n. 1, p. 44–57.

DENG, S.; HUANG, Z.; SINHA, A. P.; ZHAO, H. (2018) The Interaction Between Microblog Sentiment and Stock Returns: An Empirical Examination. **MIS Quarterly,** v. 42, n. 3, p. 895–918.

GALLIERS, R. D.; NEWEL, S.; SHANKS, G.; TOPI, H. (2017) Datification and its Human, Organizational and Societal Effects: The Strategic Opportunities and Challenges of Algorithmic Decision-Making. **Journal of Strategic Information Systems,** v. 26, n. 3, p. 185–190.

GROVER, V.; CHIANG, R. H. L.; LIANG, T.; ZHANG, D. (2018) Creating Strategic Business Value from Big Data Analytics: a research Framework. **Journal of Management Information Systems,** v. 35, n. 2, p. 388–423.

GUNTHER, W.; MEHRIZI., M.; HUYSMAN, M.; FELDBERG, F. (2017) Debating Big Data: A Literature Review on Realizing Value from Big Data. **Journal of Strategic Information Systems,** v. 26, n. 3, p. 191–209.

GUO, X.; WEI, Q.; CHEN, G.; ZHANG, J.; QIAO D. (2017) Extracting Representative Information on Intra-Organizational Blogging Platforms. **MIS Quarterly,** v. 41, n. 4, p. 1105-1127.

HAN, S.; PARK, S.; OH, W. (2016) Mobile App Analytics: A Multiple Discrete-Continuous Choice Framework. **MIS Quarterly,** v. 40, n. 4, p. 983-1008, 2016.

KELLY, S.; NOONAN, C. (2017) The doing of Datafication (and What this Doing Does). **Journal of the Association for Information Systems,** v. 18, n.12, p. 872–899.

KETTER, W.; PETERS, M.; COLLINS, J.; GUPTA, A. (2016) Competitive Benchmarking: An IS Research Approach to Address Wicked Problems with Big Data Analytics. **MIS Quarterly,** v. 40, n. 4, p. 1057–1080.

KITCHENS, B.; DOBOLYI, D.; LI, J.; ABBASI, A. (2018) Advanced Customer Analytics: Strategic Value Through Integration of Relationship-Oriented Big Data. **Journal of Management Information Systems,** v. 35, n. 2, p. 540–574.

LASH, M. T.; ZHAO, K. (2016) Early Predictions of Movie Success: The Who, What, and When of Profitability. **Journal of Management Information Systems,** v. 33, n. 3, p. 874–903.

LEHRER, C.; WIENEKE, A; BROCKE, J. V.; JUNG, R; SEIDEL, S. (2018) How Big Data Analytics Enables Service Innovation: Materiality, Affordance, and the Individualization of Service. **Journal of Management Information Systems,** v. 35, n. 2, p. 424–460.

LI, W.; CHEN, H.; NUNAMAKER, J. F. (2016) Identifying and Profiling Key Sellers in Cyber Carding Community: AZSecure Text Mining System. **Journal of Management Information Systems,** v. 33, n. 4, p. 1059–1086.

LI, T.; VAN DALEN, J.; VAN REES, P. J. (2018) More than just noise? Examining the information content of stock microblogs on financial markets. **Journal of Information Technology,** v. 33, n. 1, p. 50–69.

LOEBBECKE, C.; PICOT, A. (2015) Reflections on Societal and Business Model Transformation Arising from Digitization and Big Data Analytics: a research agenda. **Journal of Strategic Information Systems,** v. 24, n. 3, p. 149–157.

LOWRY, P. B.; DINEV, T.; WILLISON, R. (2017) Why Security and Privacy Research Lies at the Center of the Information Systems (IS) Artefact: Proposing a Bold Research Agenda. *European Journal of Information Systems,* v. 26, n. 6, p. 546–563.

LYYTINEN, K.; GROVER, V. (2017) Management Misinformation Systems: A Time to Revisit? **Journal of the Association for Information Systems,** v. 18, n. 3, p. 1–44.

LUVIZAN, S.; DINIZ, E. (2017) Big Data e o Uso Secundário de Dados: Desafios para a Qualidade de Dados e a Inovação. In: Encontro da Associação Nacional de Pós-Graduação e Pesquisa em Administração, XLI, Sao Paulo, Proceedings. Sao Paulo: ENANPAD, 2018.

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. McKinsey Global Institute.

MENON, S.; SARKAR, S. (2016) Privacy and Big Data: Scalable Approaches to Sanitize Large Transactional Databases for Sharing. **MIS Quarterly,** v. 40, n. 4), p. 963-981.

MCAFEE, A.; BRYNJOLFSSON, E. (2012) Big Data: The Management Revolution. **Harvard Business Review,** p. 1–9.

MULLER, O.; JUNGLAS, I.; BROCKE., J.; DEBORTOLI, S. (2016) Utilizing Big Data Analytics for Information Systems Research: Challenges, Promises and Guidelines. **European Journal of Information Systems,** v. 25, n. 4, p. 289–302.

MULLER, O.; FAY, M.; VOM BROCKE, J. (2018) The Effect of Big Data Analytics on Firm Performance: An Econometric Analysis Considering Temporal Dynamics and Industry Characteristics. **Journal of Management Information Systems,** v. 35, n. 2, p. 488–509.

NUAIMI, E; NEYADI, H.; MOHAMED, N; AL-JAROODI, J. (2015) Applications of big data to smart cities. **Journal of Internet Services and Applications,** v. 6, n. 25, p. 1–15.

NEWELL, S.; MARABELLI, M. (2015) Strategic Opportunities (and Challenges) of Algorithmic Decision-Making: A Call for Action on the Long-Term Societal Effects of 'Datification'. **Journal of Strategic Information Systems,** v. 24, n. 1, p. 3–14.

PARK, S.; HUH, S.; OH, W.; HAN, S.P. (2012) A Social Network-Based Inference Model for Validating Customer Profile Data. **MIS Quarterly,** v. 36, n. 4, p. 1217–1237.

1221

RAI, A. (2016) Synergies Between Big Data and Theory. **MIS Quarterly,** v. 40, n. 2, p. iii–ix.

SABOO, A. R.; KUMAR, V.; PARK, I. (2016) Using Big Data to Model Time-Varying Effects for Marketing Resource (Re) Allocation. **MIS Quarterly,** v. 40, n. 4, p. 911–939.

SHI, Z.; LEE, G.; WHINSTON, A. (2016) Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence. **MIS Quarterly,** v. 40, n. 4, p. 1035-1056.

TALLON, P.; RAMIREZ, R.; SHORT, J. (2014) The Information Artifact in IT Governance: Toward a Theory of Information Governance. **Journal of Management Information Systems,** v. 30, n. 3, p. 141–177.

The Economist (2018) Does China's digital police state have echoes in the West? Special Report on Leaders, May 31st. Acessed in 12/07/2018.
<https://www.economist.com/leaders/2018/05/31/does-chinas-digital-police-state-have-echoes-in-the-west>

The New York Times (2019) San Francisco Bans Facial Recognition Technology. Acessed in 10/06/2019.

<https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html>

WEBSTER, J.; WATSON, R.T. (2002) Analyzing Past to Prepare for Future: Writing Literature Review*,* **MIS Quarterly,** v. 26, n. 2, p. xiii—xxiii.

ZHOU, S.; QIAO, Z.; DU, Q.; WANG, G. A.; FAN, W.; YAN, X. (2018) Measuring Customer Agility from Online Reviews Using Big Data Text Analytics. **Journal of Management Information Systems,** v. 35, n. 2, p. 510–539.

ZUBOFF, S. (2015) Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. **Journal of Information Technology,** v. 30, p. 75–89.