Big data analytics in E-commerce: a systematic review and agenda for future research

2 authors:

Shahriar Akter
University of Wollongong

35 PUBLICATIONS **289** CITATIONS

Samuel Fosso Wamba
NEOMA Business School

91 PUBLICATIONS **798** CITATIONS

Abstract:

There has been an increasing emphasis on big data analytics (BDA) in e-commerce in recent years. However, it remains poorly-explored as a concept, which obstructs its theoretical and practical development. This position paper explores BDA in e-commerce by drawing on a systematic review of the literature. The paper presents an interpretive framework that explores the definitional aspects, distinctive characteristics, types, business value and challenges of BDA in the e-commerce landscape. The paper also triggers broader discussions regarding future research challenges and opportunities in theory and practice. Overall, the findings of the study synthesize diverse BDA concepts (e.g., definition of big data, types, nature, business value and relevant theories) that provide deeper insights along the cross-cutting analytics applications in e-commerce.

1. Introduction

In the past few years, an explosion of interest in big data has occurred from both academia and the e-commerce industry. This explosion is driven by the fact that e-commerce firms that inject big data analytics (BDA) into their value chain experience 5–6% higher productivity than their competitors (McAfee and Brynjolfsson, 2012). A recent study by BSA Software Alliance in the United States (USA) indicates that BDA contributes to 10% or more of the growth for 56% of firms (Columbus, 2014). Therefore, 91% of Fortune 1000 companies are investing in BDA projects, an 85% increase from the previous year (Kiron et al., 2014a). While the use of emerging internet-based technologies provides e-commerce firms with transformative benefits (e.g., real-time customer service, dynamic pricing, personalized offers or improved interaction) (Riggins, 1999), BDA can further solidify these impacts by enabling informed decisions based on critical insights (Jao, 2013). Specifically, in the e-commerce context, "big data enables merchants to track each user's behavior and connect the dots to determine the most effective ways to convert one-time customers into repeat buyers" (Jao, 2013,p.1). Big data analytics (BDA) enables e-commerce firms to use data more efficiently, drive a higher conversion rate, improve decision making and empower customers (Miller, 2013). From the perspective of transaction cost theory in e-commerce (Devaraj et al., 2002; Williamson, 1981), BDA can benefit online firms by improving market transaction cost efficiency (e.g., buyer-seller interaction online), managerial transaction cost efficiency (e.g., process efficiency-recommendation algorithms by Amazon) and time cost efficiency (e.g., searching, bargaining and after sale monitoring). Drawing on the resource-based view (RBV)(Barney, 1991), we argued that BDA is a distinctive competence of the high-performance business process to

support business needs, such as identifying loyal and profitable customers, determining the optimal price, detecting quality problems, or deciding the lowest possible level of inventory (Davenport and Harris, 2007a). In addition to the RBV, this research views BDA from the relational ontology of sociomaterialism perspective, which puts forward the argument that different organizational capabilities (e.g., management, technology and talent) are constitutively entangled (Orlikowski, 2007) and mutually supportive (Barton and Court, 2012) in achieving firm performance. Finally, service marketing offers the perspective of improving service innovation models, which has been reflected by firms such as Rolls Royce (Barrett et al., 2015), Amazon, Google and Netflix (Davenport and Harris, 2007a) . As such, the extant literature identifies BDA as the platform for "growth of employment, increased productivity, and increased consumer surplus" (Loebbecke and Picot, 2015, p.152), the "next big thing in innovation" (Gobble, 2013, p.64); "the fourth paradigm of science" (Strawn, (2012); "the next frontier for innovation, competition, and productivity" (p. 1) and the next "management revolution" (p. 3) (McAfee and Brynjolfsson, 2012); or that BDA is "bringing a revolution in science and technology" (Ann Keller et al., 2012); etc. Due to the high impact in e-commerce, notably in generating business value, BDA has recently become the focus of academic and industry investigation (Fosso Wamba et al., 2015c). As shown on Table 1, there is a steady growth in the BDA market, and in the number of global e-commerce customers and their per capita sales.

Although an increasing amount of published materials has focused on practitioners in this domain, the literature remains largely anecdotal and fragmented. There is a paucity of research that provides a general taxonomy from which to explore the dimensions and applications of big data in e-commerce. The purpose of this research therefore is to identify different conceptual dimensions of big data in e-commerce and their relevance to business value. This paper focuses

on e-commerce firms that capture business value through using big data analytics (BDA). The extant literature shows that BDA could allow an e-commerce firm to achieve a range of benefits, such as: enhanced pricing strategies for products and services (Christian, 2013); targeted advertising; better communication between research and development (R&D) and product development; improved customer service; improved multi-channel integration and coordination (IBM, 2013); enhanced global sourcing from multiple business units and locations, and, overall, suggesting models and ways to capture greater business value (Beath et al., 2012; Fosso Wamba et al., 2015a; Sharma et al., 2014). The research question that has driven this study focuses on: how is "big data analytics" different from traditional analytics in the e-commerce environment in creating business value? To answer this research question, the paper aims to provide a general taxonomy to broaden the understanding of BDA and its role in creating business value. More specifically, the aims of this paper are:

- To identify definitional perspectives of big data analytics

- To distinguish the characteristics of big data within e-commerce

- To explore the types of big data within e-commerce

- To illustrate the business value of big data in e-commerce

- To provide guidelines for tackling the challenges of big data application within e-commerce.

Overall, this paper intends to provide a thorough representation of the meaning of big data in the e-commerce context. We have organized this paper into five main parts. Firstly, in section 2, we explain the methodological gestalt and present the results of our systematic review. By collating this information, in section 3, we then define the role of big data in e-commerce and identify

alternative definitional perspectives. Secondly, in section 4, we analyze the distinctive attributes and types of big data within e-commerce. Thirdly, in section 5, we recommend different types of business value that can be derived using BDA in the e-commerce domain. Finally, we identify the challenges and provide solutions to tackle them in order to foster the growth of BDA in e-commerce.

Table 1: Global growth in e-commerce and big data analytics (BDA)

Year	Growth in the number of e-commerce customers worldwide (in millions)	Growth in e-commerce sales per customer worldwide (in US$)	Growth in big data analytics (BDA) market worldwide (in billions)
2011	792.6	1162	7.3
2012	903.6	1243	11.8
2013	1015.8	1318	18.6
2014	1124.3	1399	28.5
2015	1228.5	1459	38.4
2016	1321.4	1513	45.3

Source: Adapted from emarketer (2013) and (Piatetsky, 2014)

2. Research approach

The study was grounded in a literature review to identify and appraise the current knowledge on the definitional aspects, attributes, types and business value of BDA in e-commerce. In defining e-commerce, Kalakota and Whinston (1997) focused on four perspectives: online buying and selling, technology driven business process, communication of information and customer service. However, this definition does not provide adequate focus on transaction cost and other aspects of e-commerce (e.g., B2B, B2G, C2C etc.) Thus, illuminating these critical aspects, Frost and Strauss (2013) extends the definition focusing on buying and selling online, digital value

creation, virtual market places and storefronts and new distribution intermediaries. However, this definition heavily focuses on e-marketing and fails to integrate other important e-business processes. As such, this study puts forward a more holistic definition of e-commerce in big data environment, which aims to achieve both transaction value (i.e., cost savings, improving productivity and efficiency) and strategic value (i.e., competitive advantages, firm performance) in digital markets by transforming production, inventory, innovation, risk, finance, knowledge, relationship and human resource management with the help of analytics driven insights (Wixom et al., 2013).

This study explores 'big data' in e-commerce environment, which refers to the huge quantities of transaction, click-stream, voice, and video data in the e-commerce landscape (Davenport et al., 2012). The study embraced a systematic approach to establish rigor throughout the review: this was based on a similar approach used by Ngai and Wat (2002) and Vaithianathan (2010) in e-commerce research and Benedettini and Neely (2012) in service systems research. The review process adopted a protocol that described the criteria, scope, and methodology at each step. Due to the subjective nature of the study, the systematic approach was adapted to the specific objectives of the study. The study applied a scientific and transparent process throughout the protocol in order to make the review process more precise and less biased.

The review process was driven by the following research question: what are the definitional perspectives, distinctive characteristics, types, business value aspects and challenges of BDA in e-commerce? These aspects of the research question guided the review process by correctly identifying the subject areas, relevant studies, sources of materials, and the inclusion and exclusion criteria. The review aimed to provide pragmatic solutions to the research question by

capturing concrete and meaningful aspects with the support of empirical evidence. Therefore, major components of e-commerce (e.g., product development; operations; marketing, finance, and human resources management; and information systems) were studied in relation to BDA and business value. We have excluded disciplines which are not directly linked to our research interests, such as biology, chemistry, geology, physics, or politics. As research on BDA is an emerging field, a search within the time frame from 2006–2014 was considered to be representative. We set the lower boundary at 2006 because in this year the first seminal paper "competing on analytics" was published by Davenport (2006) in the Harvard Business Review (cited >500 times). The systematic review of the study identified this paper as a trigger for subsequent big data analytics research.

The study identified relevant publications by forming search strings that combined the key words 'big data analytics*' with a different range of terms and phrases. Using wildcard symbols, the study reduced the number of search strings because, for example, 'big data analytics*' could return hits for 'big data analytic' and 'big data analysis'. Initially, the search focused on e-commerce research as the source of material most relevant to the big data and analytics experienced by e-commerce firms. The study conducted the database search combining the key words 'big data analytics' with the terms 'electronic commerce*', 'e-commerce*', 'big data analytics* AND e-commerce*', and 'big data analytics* AND electronic commerce*'. Overall, the study identified and submitted a total of 33 search strings to a panel of experts (n=10) from different streams within e-commerce studies (i.e., marketing, operations management, strategic management, human resources management, e- commerce and information systems). The panel

represented a team of experts consisting of an academic and an analytics practitioner from each stream to validate the review protocol.

Our search started on November 01, 2014 and ended on December 20, 2014. The study reviewed scholarly peer reviewed journals, periodicals, and quality web content by exploring five databases: Scopus (Elsevier); Web of Knowledge (Thomson ISI); ABI/Inform Complete (ProQuest); Business Source Complete (EBSCO Host); and Emerald, IEEE Xplore and ScienceDirect (Taylor & Francis). In addition, a similar search was conducted within the Association of Information Systems (AIS) basket of top journals. By adding the basket of top journals, the study incorporated the key databases used by prior studies that had a similar approach (Fosso Wamba et al., 2013; Lim et al., 2013b; Ngai et al., 2008; Ngai et al., 2009) as well as important findings from leading information systems (IS) journals.

The searches were limited to the abstract field, with the exception of the Web of Knowledge database in which the topic (i.e., abstract, title, and key words) was used. A total of 121 papers were downloaded and reviewed. As the study's focus was on capturing the maximum number of views on BDA in e-commerce, a quality appraisal was established regarding the clarity of the papers' contributions to the research questions (Birnik and Bowman, 2007). At this stage, 32 papers were identified. A further seven papers were deemed relevant as they were clearly focused on BDA in various sectors including e-commerce. Cross-referencing yielded five more papers that were suitable for inclusion. At this stage, the study manually included four more papers, yielding the final list of 48 papers. Overall, the criteria used to select each paper contained an explicit or implicit indication of BDA in the e-commerce landscape.

We adopted a thematic analysis of the literature review (Ezzy, 2002) which was particularly guided by Braun and Clarke (2006). The extensive review generated a set of 5 initial codes. Although the open coding was informed by the codes derived from the literature review, the coders were open to identify additional dimensions not currently present in the literature (Spiggle, 1994). However, the coders established an initial correspondence between the literature and the five identified categories (i.e., needs identification, market segmentation, decision making and performance improvement, new product/market/business model innovation and creating infrastructure and transparency). At this stage, to establish further rigor in content analysis, we estimated Krippendorff's alpha (or, Kalpha), which is a robust reliability measure irrespective of the number of observers, levels of measurement, sample sizes, and presence or absence of missing data (Krippendorff, 2004, 2007). To estimate the Kalpha, first, each of the subsamples of 48 articles were coded independently by two judges using a nominal scale ranging from 1 to 5 (i.e., 1= needs identification, 2 = market segmentation, 3 = decision making and performance improvement, 4 = new product/market/business model innovation, 5= creating infrastructure and transparency). Second, we uploaded the coded data into IBM SPSS statistics package (version 21) and conducted the analysis following the guidelines of Hayes (2011) and De Swert (2012) to judge the inter-rater reliability of the coded variables (Hayes and Krippendorff, 2007). Finally, the results provided us a decent Kalpha value of 0.82 which exceeds the cut off value of 0.80 (De Swert, 2012), and provides us an adequate evidence of reliability in content analysis.

The overall distribution of literature that covers 5 aspects of BDA in e-commerce is presented in Table 2. It is worth noting that many articles appeared more than once as they captured multiple aspects. Clearly, the vast majority of the publications were in the category of 'decision making

and performance improvement' (48 articles or 36% of all publications). Indeed, the ultimate success of e-commerce depends on real time business decision making, which may be one explanation for the high level of publications focused on this category. The study found that most studies support the association between sophisticated analytics and robust decision making based on actionable insights. For example, using robust analytics, LinkedIn made the decision to introduce new features, such as 'People You May Know', 'Jobs You May Be Interested In', 'Groups You May Like', 'Companies You May Want to Follow' and achieved a 30% higher click-through rate (Barton and Court, 2012). This was followed by 'needs identification' and 'creating infrastructure and transparency' with 24 articles for each category (or 18% for each domain). 'Needs identification' refers to the identification of precise customer needs by exploring big transaction data, whereas 'infrastructure and transparency' focuses on making relevant information available through networks to make right time decisions. For instance, Amazon's recommendation engine generates 'you might also want' prompts to identify possible needs of customers based on the analysis of transactions history and book views (Manyika et al., 2011). As part of infrastructure and transparency, Google uses big data to refine core search and ad-serving algorithms (Davenport and Patil, 2012). Similarly, Macys, an e-retailer in the US, has developed an analytics infrastructure that can optimize pricing of 73 million items in just over one hour (Davenport et al., 2012). Macy's also analyzes data at the stock keeping unit (SKU) level to ensure that products of different assortments are readily available. Finally, the review identified 'new product/market/business model innovation' with 22 articles (or 17% of all publications), and 'market segmentation' with 14 articles or 11% of all publications. For example, Netflix Inc. created various customer segments (e.g., adventures, crime movies, family features, movies on books etc.) by analysing over one billion reviews in categories such as liked,

10

loved, hated, etc. (Davenport and Harris, 2007b). Using a visualization and demand analytics tool, Netflix introduced "House of Cards" program in the United States (USA) which is a huge success as a new product (Ramaswamy, 2013). Overall, this study identifies these five broad aspects in which organizations can create business value by harnessing big data analytics (see Table 2).

Table 2: Big data analytics (BDA) applications in e-commerce

BDA in E-commerce	References	#	%
Needs identification	(McAfee and Brynjolfsson, 2012), (Bughin et al., 2010), (Brown et al., 2011), (LaValle et al., 2011), (Allen et al., 2012), (Ann Keller et al., 2012), (Beath et al., 2012), (Boja et al., 2012), (Boyd and Crawford, 2012), (Hsinchun et al., 2012), (Demirkan and Delen, 2012), (Fisher et al., 2012), (Huwe, 2012), (Wagner, 2012),(Davenport et al., 2012), (Johnson, 2012b), (Tankard, 2012), (White, 2012b), (Bennett et al., 2013), (Minelli et al., 2013), (Lim et al., 2013a), (Constantiou and Kallinikos, 2014), (Agarwal and Dhar, 2014), (Wixom et al., 2013)	24	18%
Market segmentation	(Bughin et al., 2010), (Ann Keller et al., 2012), (Beath et al., 2012), (Boyd and Crawford, 2012), (Demirkan and Delen, 2012), (Hsinchun et al., 2012), (Griffin, 2012), (Highfield, 2012), (Smith et al., 2012), (Davenport et al., 2012), (Tankard, 2012), (Wagner, 2012), (Davenport and Harris, 2007b), (Wixom et al., 2013)	14	11%
Decision making and performance improvement	(Bughin et al., 2010), (Brown et al., 2011), (Bughin et al., 2011), (LaValle et al., 2011), (Hsinchun et al., 2012), (Boyd and Crawford, 2012), (Allen et al., 2012), (Ann Keller et al., 2012), (Boja et al., 2012), (Beath et al., 2012), (McAfee and Brynjolfsson, 2012), (Demirkan and Delen, 2012), (Davenport et al., 2012) (Fisher et al., 2012), (Gehrke, 2012), (Griffin, 2012), (Griffin and Danson, 2012), (Huwe, 2012), (Johnson, 2012b), (Johnson, 2012a), (Lane, 2012), (Ohata and Kumar, 2012), (Smith et al., 2012), (Tankard, 2012), (Wagner, 2012), (White, 2012b), (Bennett et al., 2013), (Koch, 2013), (Rajpurohit, 2013), (Mithas et al., 2013), (Kung et al., 2013), (Dimon, 2013), (Minelli et al., 2013), (Davenport, 2006), (Viaene and Van den Bunder, 2011), (Waller and Fawcett, 2013), (Goes, 2014), (George et al., 2014), (Constantiou and Kallinikos, 2014), (Agarwal and Dhar, 2014; Agarwal and Weill, 2012), (Kiron et al., 2012a), (Kiron et al., 2014a), (Mithas et al., 2013), (Wixom et al., 2013), (Mehra, 2013), (Kopp, 2013), (Miller, 2013)	48	36%
New product/market /business model innovations	(Bughin et al., 2010), (Bughin et al., 2011), (LaValle et al., 2011), (Brown et al., 2011), (Ann Keller et al., 2012), (Beath et al., 2012), (Boyd and Crawford, 2012), (McAfee and Brynjolfsson, 2012), (Hsinchun et al., 2012), (Demirkan and Delen, 2012), (Fisher et al., 2012), (Gehrke, 2012), (Griffin, 2012), (Griffin and Danson, 2012), (Huwe, 2012), (Johnson, 2012b), (Ohata and Kumar, 2012), (Tankard, 2012), (Wagner, 2012), (Kung et al., 2013), (Lim et al., 2013a), (Wixom et al., 2013)	22	17%
Creating infrastructure and transparency	(McAfee and Brynjolfsson, 2012), (Brown et al., 2011), (Bughin et al., 2011), (LaValle et al., 2011), (Ann Keller et al., 2012), (Beath et al., 2012), (Boyd and Crawford, 2012), (Hsinchun et al., 2012), (Fisher et al., 2012), (Griffin, 2012), (Huwe, 2012), (Smith et al., 2012), (Tankard, 2012), (Wagner, 2012), (Zeng and Lusch, 2013), (Barton and Court, 2012), (Waller and Fawcett, 2013), (Goes, 2014), (Constantiou and Kallinikos, 2014), (Agarwal and Dhar, 2014), (Kiron et al., 2014a), (Mithas et al., 2013), (Wixom et al., 2013), (Leavitt, 2013)	24	18%
	Total	132	100%

Note: Some articles are counted more than once because they cover more than one type of BDA in e-commerce.

3. Defining big data analytics in the e-commerce environment

E-commerce firms are one of the fastest groups of BDA adopters due to their need to stay on top of their game (Koirala, 2012). In most cases, e-commerce firms deal with both structured and unstructured data. Whereas structured data focuses on demographic data including name, age, gender, date of birth, address, and preferences, unstructured data includes clicks, likes, links, tweets, voices, etc. In the BDA environment, the challenge is to deal with both types of data in order to generate meaningful insights to increase conversions. Schroeck et al. (2012) found that the definition of big data incorporated various dimensions including: greater scope of information; new kinds of data and analysis; real-time information; non-traditional forms of media data; new technology-driven data; a large volume of data; the latest buzz word; and social media data. In defining big data, IBM (2012), Johnson (2012a), and Davenport et al. (2012) focused more on the variety of data sources, while other authors, such as Rouse (2011), Fisher et al. (2012), Havens et al. (2012), and Jacobs (2009) emphasized the storage and analysis requirements of dealing with big data. As defined by IDC (2013), big data focuses on three main characteristics: the data itself, the analytics of the data, and the presentation of the results of the analytics that allow the creation of business value in terms of new products or services. Overall, the study defines big data in terms of five Vs: volume, velocity, variety, veracity, and value (White, 2012). The 'volume' refers to the quantities of big data, which is increasing exponentially. The 'velocity' is the speed of data collection, processing and analyzing in the real time. The 'variety' refers to the different types of data collected in big data environment. The 'veracity' represents the reliability of data sources. And, finally, the 'value' represents the

transactional, strategic and informational benefits of big data (Fosso Wamba et al., 2015b; Wixom et al., 2013).

The sheer volume of academic and industry research provides evidence on the importance of big data in many functional areas of e-commerce including marketing, human resources management, production and operation, and finance (Agarwal and Weill, 2012; Bose, 2009; Davenport, 2006; Davenport, 2010, 2012; Davenport et al., 2012). In e-commerce, a large amount of customer-related information is available simply when customers 'sign in': these data are of great interest to business decision makers. While the significance of big data in making strategic decisions is recognized and understood, there is still a lack of consensus on the operational definition of big data analytics (BDA). It is thus prudent to analyze the definitions of BDA mentioned in previous studies in order to identify their common themes. For example, Davenport (2006) indicated that BDA refers to the quantitative analysis of big data with a view to making business decisions. In addition, this decision-usefulness aspect of analytics has been the focus in other studies such as those by Davenport and Harris (2007b), Davenport (2010), and Bose (2009). Whereas Davenport and Harris (2007b) explained BDA with the help of mechanisms such as statistical analysis and the use of an explanatory and predicting model, Bose (2009, p.156) described BDA as the "group of tools" used to extract, interpret information as well as predict the outcomes of decisions.

In defining BDA, one stream of research has focused on strategy-led analytics, or analytics that create sustainable value for business. For example, LaValle et al. (2011) explained that the application of business analytics (or the ability to use big data) for decision making must essentially be connected with the organization's strategy. Indeed, strategy-driven analytics has

14

received much attention due to its role in better decision making. Studies have also focused on "competitive advantages" and "differentiation", while applying analytics to analyze real-time data (Schroeck et al., 2012). In a similar vein, Biesdorf et al. (2013) highlighted that it is important to create an environment where big data, process optimization, frontline tools and people are well-aligned in order to achieve competitive advantages.

The other stream of research defines BDA from the perspective of identifying new opportunities with big data (see Appendix 1). For example, Davenport (2012) explained that BDA attempts to explore new products and value-added activities. Similar arguments have also been offered in another study by Davenport et al. (2012) in scanning the external environment and identifying emerging events and opportunities. In addition, studies have also highlighted the role of behavioral elements in the BDA definition (Agarwal and Weill, 2012; Ferguson, 2012), such as empathy, which they believe to be essential in considerably enhancing the analytical ability of firms. They explained BDA as being a combination of three things, that is, business processes, technology optimization, and emotional connection with the use of data. In a similar spirit, Ferguson (2012) stated that BDA refers to a multidimensional behavioral analysis that covers both internal and external factors.

Overall, it is evident that statistical, contextual, quantitative, predictive, cognitive, and other models are necessary prerequisites for big data analytics (BDA) (Kiron et al., 2012a). As such, the study defines BDA in e-commerce as a holistic process that involves the collection, analysis, use, and interpretation of data for various functional divisions with a view to gaining actionable insights, creating business value, and establishing competitive advantage. While this definition reflects the notion that analytical techniques can be used to produce actionable insights is rooted

in the origin of statistics traced back to the 18th century, the clear difference today is the vast amount electronic transactions in the digital economy and its associated data deluge(Agarwal and Dhar, 2014). From the transaction cost theory and the new institutional economics viewpoint (Williamson, 1979, 1981, 2000), when it comes to economic performance of e-commerce firms in the emerging data economy, institutional structure can play pivotal role in defining BDA. In addition, relationship based e-commerce theories in terms of trust, loyalty and privacy (Dinev and Hart, 2006; Gefen, 2000, 2002) or, classic information systems theories, such as IS success (Delone, 2003; DeLone and McLean, 1992), IT quality (Wixom and Todd, 2005), IS continuance (Bhattacherjee, 2001) and IT capability, sociomateriality and business value theories (Kim et al., 2012) can be utilized in defining BDA based on the objectives and scope of the study. As a result, there are clear opportunities for theoretical and practical inquiry to define BDA in a particular e-commerce context ranging from marketing promotions, customer relationship to supply chain management. By developing interesting research questions, this new frontier of data science can create new knowledge and scientific possibilities by leveraging on data, technology, analytics, business and society. Appendix 1 summarizes the definitional aspects and potential research areas on BDA in e-commerce.

4. Big data and their distinctive characteristics in the e-commerce environment

The e-commerce landscape today is bubbling up with numerous big data that are being used to solve business problems. According to Kauffman et al. (2012, p.85), the use of big data is skyrocketing in e-commerce "due [to] the social networking, the internet, mobile telephony and all kinds of new technologies that create and capture data". With the help of cost-effective storage and processing capacity, and cutting-edge analytical tools, big data now enable e-commerce firms to reduce costs and generate benefits without any difficulty. However, analytics that capture big data is different from traditional data in many respects. Specifically, owing to the elements of their unique nature (i.e., voluminous, variety, velocity, and veracity), big data can be easily distinguished from the traditional form of data used in analytics (see Appendix 2). The next sections discuss these elements in turn, along with their implications for e-commerce.

4.1 Voluminous

With the emergence of web technologies, there is an ever-increasing growth in the amount of big data in the e-commerce environment (Beath et al., 2012). These mass quantities of data that e-commerce firms are trying to harness to improve their decision-making process are defined as voluminous (McAfee and Brynjolfsson, 2012). As illustrated by Russom (2011), BDA takes a large volume of data that require a massive amount of storage and entail a large number of records. In fact, BDA encompasses large volumes of data (commonly expressed in petabytes and exabytes) that are used by decision makers for making strategic decisions. Data collected in the big data environment are often unstructured and can incorporate video, image, or data generated

from mobile technology. As such, it is unlikely that big data will be clean and free from any errors. While this poses an extra challenge for decision makers to get data ready for use, big data enable real-time decision making for e-commerce firms (Kang et al., 2003). For example, using the massive amount of structured and unstructured data, Amazon developed sophisticated recommendation engines that deliver over 35 percent of all sales, automated customer service systems to ensure superior customer satisfaction and dynamic pricing systems that adjust pricing against competing sites every 15 seconds (Goff et al., 2012). Similarly, Netflix, the online movie retailer, analyzes over 1 billion reviews to determine the customer's movie tastes and inventory conditions (Davenport and Harris, 2007b). Many e-commerce firms (e.g., Amazon, eBay, Expedia, Travelocity) use massive volume of social media data (e.g., photos, notes, blog posts, web links, and news stories) to tap into the opportunity of real time promotional offers (Manyika et al., 2011). In addition to opportunities, the volume of big data brings challenges, especially integration of big data from different sources and formats, introducing new "agile" analytical methods and machine-learning techniques, and increasing the speed of data processing and analysis. As such, E-commerce firms must have the ability to embed analytics and optimization into their operational and decision processes to enhance their speed and impact (Davenport, 2013a).

4.1 Variety

The word 'variety' denotes the fact that big data originate from numerous sources which can be structured, semi-structured or unstructured (Schroeck et al., 2012). Variety is another critical attribute of big data as they are generated from a wide variety of sources and formats including text, web, tweet, audio, video, click-stream, log files, etc. (Russom, 2011). This variety of data requires the use of different analytical and predictive models which can enable information about

different functional areas to be used. Biesdorf et al. (2013) explained, for example, that the analytic model used by e-commerce firms could comprise a variety of customer information, such as: customer profiles and historical data on buying behavior; regional and seasonal buying patterns; optimizing of supply chain operations; and, above all, the retrieval of any unstructured data from social media to predict buying by product, store, and advertising activities. For example, Manyika et al. (2011) showed that an e-retailer provided real-time responses in marketing campaigns, amending them as and when necessary by conducting sentiment analysis. Overall, the variety of big data has the potential to add business value to firms. However, top management commitment in terms of improving business processes and defining workflows is very significant in order for the benefits from such data to be realized (Beath et al., 2012).

4.2 Velocity

Velocity refers to the frequency of data generation and/or the frequency of data delivery (Russom, 2011). It is important to understand the velocity of big data which needs to be prioritized and synced into business processes, decision making and improvements in performance (Beulke, 2011). As described by Gentile (2012), the term 'velocity' is the rate of change in big data and how quickly big data should be used in business decisions in order to add value. In fact, given that greater data velocity is assured, data have the potential to open up new opportunities for organizations. As shown by Davenport and Patil (2012), the high velocity of BDA can enable analysts to conduct consumer sentiment analysis and provide a clear picture about choices of brands and/or products.

To capitalize on this high pace of data, many e-commerce firms have used various techniques to add value to their business. For example, Amazon has been able to maintain a constant flow of

new products by right-time communication with its stakeholders (Davenport, 2006). eBay Inc. has performed thousands of experiments using data velocity with different aspects of its website, which has resulted in better layout and website features ranging from navigation to the size of its images (Bragge et al., 2012). To utilize the high velocity of data, many e-commerce firms now use sophisticated systems to capture, store, and analyze the data to make real-time decisions and retain their competitive advantages.

4.3 Veracity

Another essential attribute of big data relates to the uncertainty associated with certain types of data. These data demand rigorous verification, requiring full compliance with quality and security issues. High data quality is an important requirement of BDA for better predictability in the e-commerce environment (Schroeck et al., 2012). Therefore, verification is necessary to generate authenticated and relevant data, and to have the capability to screen out bad data (Beulke, 2011). In fact, verification is essential in the data management process because the existence of bad data might hinder management in making cognizant decisions. Likewise, bad data would have little relevance in adding business value. Beulke (2011) explained that the information technology (IT) units of e-businesses can play a key role in this regard by setting up an automatic verification system so that the big data used for business decisions have been authenticated and have passed through strict quality compliance procedures.

In this regard, Schroeck et al. (2012) argued for the use of data fusion which combines various less reliable data sources in order to create a more precise and worthwhile data point (such as social comments affixed to geospatial location information). Ferguson (2012) highlighted that Montage Analytics has developed a tool which is particularly useful for predicting 'black

swans[1], in organizations, and any other types of risk that have originated as a result of human manners and motivations. The reason is that the inherent unpredictability of some data is always caused by factors, such as technology failure, human lack of truthfulness and economic factors.

5. Types of big data used in e-commerce

E-commerce refers to the online transactions: selling goods and services on the internet, either in one transaction (e.g., Amazon, Zappos, eBay, Expedia) or through an ongoing transaction (e.g., Netflix, Match.com, LinkedIn etc.) (Frost and Strauss, 2013). E-commerce firms ranging from Amazon to Netflix capture various types of data (e.g., orders, baskets, visits, users, referring links, keywords, catalogues browsing, social data), which can be broadly classified into four categories: (a) transaction or business activity data (b) click-stream data (c) video data and (d) voice data (see Appendix 3). In e-commerce, data are the key to track consumer shopping behaviour to personalize offers, which are collected over time using consumer browsing and transactional points. This section discusses different types of big data along with their implications for e-commerce.

5.1 Transaction or business activity data

Transaction or business activity data evolve as a result of exchanges between the customer and company over time. These data are structured in nature and originate from many sources ranging from customer relationship programs (e.g., customer profiles maintained by the company, the

---

[1] A 'black swan' is also known as an extreme outlier: it is basically the disproportionate quality of a high-impact, hard-to-predict, rare event that, in general terms, people do not predict to happen, an event that is beyond the scope of detection by traditional analytics (Ferguson, 2012).

occurrence of customer complaints) through to sales transactions. A recent study by Chandrasekaran et al. (2013) provided the example of an e-retailer that analyzes data from its loyalty program (i.e., its Clubcard loyalty program), entailing 1.6 billion data points, 10 million customers, 50,000 stock keeping units (SKUs), and 700 stores, which has resulted in the thorough coordination of big data with consumer insights. In the context of e-retailing, Kiron et al. (2014b) reported that StyleSeek, the online recommendation engine in the US, makes massive revenue by analyzing customers' tastes and preferences and driving consumers toward its retail partners with the help a sophisticated analytics platform. Overall, it is evident that e-retailers can derive numerous benefits across the value chain using transaction data.

5.2 Click-stream data

Click-stream data originate from the web and online advertisements, and from social media content such as the tweets, blogs, Facebook wall postings, etc. of e-commerce businesses. In today's connected environment, social media and online advertisements play a key role in the ongoing promotional strategy of firms, such as the use of click-stream data that are very important for management in making informed, strategic, and tactical decisions. Earlier studies have found that many e-commerce firms worldwide (e.g., Amazon, eBay, Zappos, Alibaba etc.) rely on click-stream data in their efforts to capture data. Click-stream data can be applied to predict customer preferences and tastes. As highlighted by Davenport and Harris (2007a), Netflix, a world-famous internet TV network, captures and analyzes more than one billion web data related to reviews of movies that are liked, loved, hated, etc. in order to understand customers' tastes.

Another recent study by Davenport et al. (2012) reported that credit card firms, through relying on website and call center data, maintain databases (named as ready-to-market) to offer customer-tailored products within milliseconds and also optimize offers by following up responses from customers. Some companies use such databases not only to approach customers but also to offer online services. For example, Biesdorf et al. (2013) explained that by analyzing web data, e-retailers receive a red flag alert when the prices of their competitors' products are below their own price level. Therefore, retailers can adjust their prices to remain competitive.

5.3 Video data

Video data are live data that come from capturing live images. Currently, e-commerce firms are keen to use not only either click-stream data or transaction data but, in association with image analysis software, they tend to also capture video data. As indicated by Schroeck et al. (2012), e-commerce firms have the necessary competencies to analyze extremely unstructured data, such as video or voice data. These data have the potential to add value for e-commerce firms. For example, Ramaswamy (2013) reported that Netflix uses video data to predict viewing habits and evaluate the quality of experiences. In addition, the visualization and demand analytics tool based on the type of movie consumption help Netflix understand preferences, which led them to achieve success in their "House of Cards" program in the US. Thus, the use of video data is essential for firms in making better decisions than their competitors.

5.4 Voice data

Another type of data attached to the big data family is voice data, that is, data typically originating from phone calls, call centers, or customer service. As evidenced in recent research, voice data are advantageous for analyzing consumer-buying behavior or targeting new

customers. As described by Davenport et al. (2012), credit card companies, for example American Express, use and track data related to call center activities so that personalized offers can be given in milliseconds. In Schroeck et al.'s (2012) survey, e-commerce firms were found to use advanced capabilities to analyze text and transcripts converted from call center conversations. In addition, numerous nuances of language, such as sentiment, slang and intentions, can be read and recognized by means of BDA in the context of e-commerce.

Since the nature and type of big data are unique and coming from various networks of digital platforms, there is possibility of new theory enquiring new problems. The data economy also indicates that big data are "relational" and "networked", which necessitate new developments in IT capability and algorithms, system and data quality, privacy and ethical implications, strategic alignment and corporate culture.

6. Business value of big data analytics for e-commerce firms

The ultimate challenge of BDA is to generate business value from this explosion of big data (Beath et al., 2012). The term 'value' in the context of big data implies the generation of economically worthy insights and/or benefits by analyzing big data through extraction and transformation. Aligned with Wixom et al. (2013), we define business value of BDA as the transactional, informational and strategic benefits for the e-commerce firms. Whereas transactional value focuses on improving efficiency and cutting costs, informational value sheds light on real time decision making and strategic value deals with gaining competitive advantages. For example, by injecting analytics into e-commerce, managers could derive overall business

value by serving customer needs (79%); creating new products and services (70%); expanding into new markets (72%); and increasing sales and revenue (76%) (Columbus, 2014). Table 3 shows that many e-commerce firms worldwide are able to enhance business value in the form of transactional, informational and strategic benefits by using big data analytics.

Amazon, the online retailer giant, is a classic example of enhancing business value and firm performance using big data. Indeed, the firm was able to generate about 30% of its sales through analytics (e.g., through its recommendation engine) (*The Economist*, 2011). Similarly, Kiron et al. (2012b) reported that Match.com was able to earn over 50% increase in revenue in the past two years while the company subscriber base for its core business reached 1.8 million. The IBM case study (IBM, 2012) illustrated that greater data sharing and analytics could improve patient outcomes. For example, Premier Healthcare Alliance was able to reduce expenditure by US$2.85 billion. Schroeck et al. (2012) found that Automercados Plaza's grocery chain was able to earn a nearly 30% rise in revenue and a total of US$7 million increase in profitability each year by implementing information integration throughout the organization. Furthermore, the company avoided losses on over 30% of its products by scheduling price reductions to sell perishable products on time. In addition to adding value for business in financial terms, the use of big data can add benefit in non-financial parameters such as customer satisfaction, customer retention, or improving business processes. As presented by Davenport (2006), United Parcel Service (UPS) examines usage patterns and complaints data to accurately predict customer defections. This process has resulted in a significant increase in customer retention for the firm. In the similar spirit, LaValle et al. (2011) reported that an online automobile company was able to develop accurate customer retention strategies by creating a customer sample from big data followed by

applying analytic algorithms to forecast attrition probabilities, coupled with identifying at-risk customers. This retention strategy consequently has opened up prospects for the firm to yield hundreds of millions of dollars merely from a single brand. Because e-commerce firms have opportunities to interact in real time with customers more frequently than firms that do not have an e-commerce platform, they need to use big data for various purposes. Therefore, drawing on the relevant theories in e-commerce (see Table 4), we present six mechanisms to enhance practical business values in data economy as follows.

6.1 Personalization

The first application of big data for e-commerce firms is the provision of personalized service or customized products (Koutsabasis et al., 2008). Studies have argued that consumers typically like to shop with the same retailer using diverse channels, and that big data from these diverse channels can be personalized in real time (Kopp, 2013; Mehra, 2013; Miller, 2013). Real-time data analytics enables firms to offer personalized services comprising special content and promotions to customers. In addition, these personalized services assist firms to separate loyal customers from new customers and to make promotional offers accordingly (Mehra, 2013). According to Liebowitz (2013), personalization can increase sales by 10% or more and provide five to eight times the ROI on marketing expenditures. Bloomspot, in this regard, explored customer credit card data to track the spending records of the most loyal customers and to offer them rewards via follow-up offers and benefits which assisted in increasing customer loyalty (Miller, 2013). Wine.com achieved a massive increase in their sales using personalized email marketing (Zhao, 2013). Bikeberry.com is an example of an e-commerce firm that is now leveraging BDA (e.g., using data from customers' browsing patterns, login counts, past

purchases) to send each customer a tailored offer: this has led to a sales increase of 133% and user on-site engagement increase of approximately 200% (Jao, 2013).

6.2 Dynamic pricing

In today's extremely competitive market environment, customers are considered 'king'. Therefore, to attract new customers, e-commerce firms must be active and vibrant while setting a competitive price (Kung et al., 2013). Amazon.com's dynamic pricing system monitors competing prices and alerts Amazon every 15 seconds, which has resulted in a 35% increase in all sales. To offer competitive prices to customers on the eve of possible increases in sales (such as at Christmas or other festive times), Amazon processes big data by taking into account competitors' pricing, product sales, actions of customers, and any regional or geographical preferences (Kopp, 2013). Access to this information through the use of big data is likely to enable e-commerce firms to establish dynamic pricing (Leloup and Deveaux, 2001).

6.3 Customer service

Another key area in which e-commerce firms can use big data is customer service. Customer grievances communicated by means of contact forms in online stores together with tweeting enable e-commerce firms to make customers feel valued when they call the service center resulting in prompt service delivery (Mehra, 2013). Similarly, Miller (Miller, 2013) explained that, by offering proactive maintenance (i.e., taking preventive measures before a failure takes place or is even detected) using big data obtained from sensors rooted in products, e-commerce firms are able to offer innovative after-sales service.

6.4 Supply chain visibility

When customers place an order on an online platform, it is logical for them to expect that companies would provide the service of tracking the order while the goods are still in transit. Kopp (2013) explained that customers expect key information, such as the exact availability, current status, and location of their orders. E-commerce firms often face difficulty in addressing these expectations from customers as various third parties such as warehousing and transportation are involved in the supply chain process (Kopp, 2013). Big data analytics (BDA) plays a key role in this context by collecting multiple information from multiple parties on multiple products (Mehra, 2013), and subsequently precisely advises the expected delivery date to customers.

6.5 Security and fraud detection

Fraud-related losses, on average, amount to US$9,000 for every US$1 million in revenue (Mehra, 2013). This significant amount of loss can be prevented by identifying relevant insights through the use of big data. With the help of the right infrastructure, such as Hadoop, e-commerce firms can analyze data at an aggregated level to identify fraud relating to credit cards, product returns and identity theft (Mehra, 2013). In addition, e-commerce firms are able to identify fraud in real time by combining transaction data with customers' purchase history, web logs, social feed, and geospatial location data from smartphone apps. For example, Visa has installed a big data-enabled fraud management system that allows the inspection of 500 different aspects of a transaction, with this system saving US$2 billion in potential losses annually.

6.6 Predictive analytics

Predictive analytics refers to the identification of events before they take place through the use of big data (Kopp, 2013). The application of predictive analytics depends on robust data mining (Cherif and Grant, 2013). In this context, Loveman (2003), CEO and President of Caesar's Entertainment, stated that: "[t]he best way to engage in … data-driven marketing is to gather more and more specific information about customer preferences, run experiments and analyses on the new data, and determine ways of appealing to [casino game] players' interests. We realized that the information in our database, coupled with decision science tools that enabled us to predict individual customer's theoretical value to us, would allow us to create marketing interventions that profitably addressed players' unique preferences." Therefore, predictive analytics helps firms to prepare their revenue budgets. The preparation of these budgets assists e-commerce firms to recognize future sales patterns from past sales data (e.g., yearly or quarterly). This, in turn, helps firms to better forecast and determine inventory requirements, thus leading to the avoidance of product stockouts and lost customers (Mehra, 2013). Similarly, the application of a visualization and demand analytics tool at Netflix helped in the accurate prediction of consumer behavior and preferences when airing the "House of Cards" program in the United States (USA) (Ramaswamy, 2013).

E-commerce firms increasingly extract business value from BDA insights either to solve business problems or make decisions. This new development in the realm of data driven e-commerce triggers development of new theories in the context of tangible (e.g., productivity improvement) and intangible (e.g., strategic business understanding) business value using people, process and technology. For example, Wixom et al. (2013) identified 'drivers of pervasive use' and 'drivers of speed to insight' as new constructs for possible theoretical

extension in BDA adoption and value. In the similar spirit, Fosso Wamba et al. (2015a) highlighted transactional, strategic and transformative business values of BDA as fertile grounds for knowledge creation. Also, Sharma et al. (2014) put forward new research questions exploring the roles of resource allocation and resource orchestration processes in organizations in order to gain a deeper understanding of how BDA influences business value. Similar arguments have been put forward by Beath et al. (2012) to explore BDA and business value in practice to capitalize on the current opportunities of big data. Overall, we suggest conducting innovative, non-traditional research in electronic markets that leverages the power of big data towards theory building in the form of capturing novel social and economic activity.

Table 3: Business value of big data analytics (BDA) in e-commerce

Study	E-Commerce Functions	Description	Firm(s)
Davenport (2006)	Personalization (e.g., CRM)	Identify customers with the greatest profit potential, loyalty, and service. Increase likelihood that they will want the product or service offering, retain their loyalty.	Harrah's, Capital One, Barclays
	Dynamic pricing	Identify the price that will maximize yield or profit.	Progressive, Marriott
	Customer service (e.g., service quality)	Detect quality problems early and minimize them.	Honda, Intel
	Predictive analytics	Fine tuning global promotions for every medium in every region.	Dell (DDB Matrix)
	Supply chain visibility	Analysts from such functions as operations and supply chain to improve total business performance by analyzing interrelationships among functional areas.	Procter & Gamble (P&G)
	Predictive analytics (e.g., churn)	Customer Intelligence Group examines usage patterns and complaints data to accurately predict customer defections.	United Parcel Service (UPS)
Manyika et al. (2011)	Predictive analytics (e.g., market share analysis)	Use of big data to capture market share from local competitors.	Tesco
	Personalization (e.g., direct marketing, relationship marketing)	Recommendation engine to generate 'you might also want' prompts to generate sales.	Amazon.com
	Predictive and cluster analytics (e.g., segmentation, customer profitability).	Developed behavioral segmentation and a multi-tier membership reward program by analyzing customer profiles, real-time changes in customer behavior, and customer profitability.	Neiman Marcus
	Personalization (email marketing)	Integrated customer databases with information on some 60 million households to improve the response rate of email marketing.	Williams-Sonoma
	Personalization (customized service offerings)	Compiled detailed holistic customer profiles, and conduct experiments and segment their customers systematically and effectively to personalize product offers and increase customer loyalty.	Harrah's, Progressive Insurance, Capital One
(Davenport and Harris, 2007b)	Dynamic Pricing	Deriving the most accurate pricing of products and services with precise calculation of customer profitability.	Royal Bank of Canada

	Customer choice preferences and product offerings	Analyzes customer choice and customer feedback from over one billion reviews.	Netflix
(LaValle et al., 2011)	Predictive analytics (data-driven customer insights)	Collected 80–90% of the information possibly needed about customers to generate analytics-driven customer insights.	Best Buy
(Davenport et al., 2012)	Dynamic pricing	Is able to optimize pricing of 73 million items in just over one hour.	Macys.com
(Kiron et al., 2012a)	Customer service (e.g., service innovation)	Uses personal profiles and psychology-based analytics to help people connect and enter into a loving relationship.	Match.com
	Security and fraud detection	Each new PayPal initiative across finance, operations, and products is examined with quantified impact and leveraging analytics.	PayPal
(Schroeck et al., 2012)	Dynamic pricing	Scheduling price reductions to sell perishable products before they spoil.	Automercados Plaza's
(Chandrasekaran et al., 2013)	Cluster and predictive analytics (segmentation)	Systematically integrates analytics and consumer insights using data from its Clubcard loyalty program to better segment and target customer occasions.	Tesco
	New product acceptance rate	Simulates new products placed on shelves in order to test design effects internally and with consumers in order to enhance product acceptability after launching.	Procter & Gamble (P&G)
(Davenport and Patil, 2012)	(a) core search (b) advertisements	Uses BDA to refine core search and ad-serving algorithms.	Google
	Product, feature (e.g., 'people you may know') and value-adding service	Uses BDA to generate ideas for products, features, and value-added services. By using 'people you may know', LinkedIn generated millions of new page views which resulted in LinkedIn's growth trajectory shifting significantly upward.	LinkedIn
(Liebowitz, 2013)	Product management	Macy's analyzes data at the stock keeping unit (SKU) level to ensure that it has readily available assortments of products.	Macys.com

7. Discussion and future research agenda

While the use of big data tends to add value for business throughout the entire value chain, there are a few challenges that organizations should confront and resolve before pay-offs from big data will flow into their business. Indeed, any innovative way of performing jobs always brings challenges: big data is no exception to this reality. Many researchers have argued that, while big data have great potential to improve business performance/add value, decision makers need to address various business challenges in order to reap the benefits (Davenport, 2012; Schroeck et al., 2012; Shah et al., 2012). As shown in Table 4, the current study highlights some of these challenges with theoretical and practical implications, thus laying the ground for potential research on BDA in the e-commerce landscape.

One of the biggest challenges in the big data environment is that it does not give clear direction on how to reach business targets by aligning with the existing organizational culture and capabilities (Kiron et al., 2014a). In this regard, Barton (2012) highlighted that the key challenge for managers is to make big data trustworthy and understandable to frontline employees, with the example that frontline employees are typically reluctant to use big data as they either did not trust a big data-based model or did not have the capabilities to understand how it worked. Therefore, in the process of gaining greater acceptance by employees and other end-users, managers should present big data in an understandable format such as through a dashboard, reports or a visualization system (Bose, 2009). Indeed, an innovative capability always leads toward sustained long-term advantages (Porter and Millar, 1985) and superior firm performance through the characteristics of rarity, appropriability, non-reproducibility, and non-substitutability

(Barney, 1991). Therefore, the required training, discussion with relevant employees and managers, the active role of top management (a leader), and incentives could work as catalysts to facilitate the adoption by employees and managers of big data (Kiron et al., 2014a). In this regard, McAfee and Brynjolfsson (2012) argued that it is very unlikely for a firm to be a top performer through the use of big data unless there is a clear goal and strategy in place.

Marketing within e-commerce firms is grappling with the massive amount of data arriving through multiple channels. Therefore, the biggest challenge is to find the right information about each customer from the large amount of data (Agarwal and Dhar, 2014; Miller, 2013). This huge amount of data empowers customers more than ever before and creates an opportunity for marketers to establish relationship in e-commerce based on trust and loyalty (Gefen, 2000, 2002). Although organizations such as Facebook, Google, and Twitter have an enormous amount of person-specific information, the challenge is how they and other e-commerce firms can inject BDA into marketing practices to create personalized offers, set dynamic prices, and use the right channels to provide consumer value. In addition to marketing, it is also important to process big data in production and operations management using sophisticated database management tools. In relation to e-commerce, big data advocates argue for the integration of talent, technology and information to reduce overall transaction costs (Williamson, 1979, 1981). In this regard, Chang et al. (2014, p.12) stated that: "[s]ince big data are now everywhere and most firms can acquire it, the key to competitive advantage is to accelerate managerial decision-making by providing managers with implementable guidelines for the application of data analytics skills in their business processes".

Table 4: Future research questions for big data analytics (BDA) in e-commerce

E-Commerce research streams	Relevant theories	Future Research Questions for BDA in E-Commerce
Strategy, culture, leadership, and organization	Resource based theory (Barney, 1991), Competitive strategy (Porter and Millar, 1985)	• How can organizations better incorporate functional differences into their big data strategy in order to develop a big data-oriented culture? • How can organizations ensure business alignment, business process transformation, and strategic analytics in the big data environment?
Marketing and sales	Market orientation theory (Kohli and Jaworski, 1990; Narver and Slater, 1990), relationship theory in e-commerce (Gefen, 2000, 2002), consumer value (Holbrook, 1999)	• What factors influence the consumer-perceived value of big data? • What offerings generate more profit and which segments are more attractive? • How can pricing be customized for the individual customer? • Factors influencing trust and loyalty of customers in e-commerce?
Production and operations management	Transaction cost theory (Williamson, 1979, 1981)	• What is the impact of big data on lean operations, quality management, and supply chain management? • How can organizations better use insights from big data to achieve operational excellence? • What is the impact of big data on the operations of e-commerce in various sectors (e.g., healthcare, retail industry, and manufacturing)?
Data quality, IT infrastructure, and security	IT quality theory (Nelson et al., 2005), IS success theory (Delone, 2003; DeLone and McLean, 1992), Sociomateriality of IT (Orlikowski, 2007)	• What factors influence data cost, quality, retention, visualization, governance, security, and privacy? • How can an e-commerce firm combine all channels to generate, acquire, transform, and integrate big data? • What types of contracts can allow firms to use big data for legitimate purposes? • What are the BDA capability dimensions for e-commerce firms? • How can a firm leverage big data sources and cloud-based architecture to produce insight and business value? • Do privacy and security concerns have any impact on customers?
Human resources/talent management	Resource based theory (Barney, 1991)	• What factors influence the selection and retention of data scientists? • What types of training improve the skills and engagement of employees at all levels to deal with big data?
Overarching value	IT business value (Melville et al., 2004), business value of analytics (Wixom et al., 2013)	• How do business analytics and organizational decision making process influence each other and what are their joint effects on firm performance? • How does big data adoption and implementation vary by firm types in the e-commerce industry? • How do organizations capitalize on big data at their disposal and extract value? • What factors influence the data life cycle in the e-commerce industry? What are their implications at different stages?

The availability of good quality big data is the key to adding value to the organization (Kiron et al., 2014a). Poor quality data might arise from redundant applications and databases, which add to data storage costs and make data more difficult to access and use (Beath et al., 2012). Although big data can be leveraged to improve business value, there is always the risk of redundant, inaccurate, and duplicate data which might undermine he decision-making process (Nelson et al., 2005). As argued by Schroeck et al. (2012), poor data quality or ineffective data governance is a key challenge for BDA. It is noteworthy that the use of even the most sophisticated analytics would be meaningless if inappropriate data are in place or poor quality data are used (Bose, 2009).

In addition to good quality data, the safe handling of individual and organizational privacy and data security (e.g., names and addresses, social security numbers, credit card numbers, and financial information) could possibly be another challenge for big data management (Bose, 2009; Smith and Shao, 2007). Although the unprecedented growth in BDA makes it alluring to use data without consent, that non-consenting respondents might jeopardise the advancement in research. As such, informed consent is a big challenge in big data environment (Bialobrzeski et al., 2012). Indeed, informed consent process need to be streamlined in BDA by embracing flexible, refined, simplified but informative consent process (Beskow et al., 2010; Ioannidis, 2013; White, 2012a) to encourage participatory community research (Bouhaddou et al., 2011). Undoubtedly, big data provides actionable insights for e-commerce firms, however, it creates a "privacy paradox" because consumers, on the one hand, want to protect their privacy and on the other hand, they regularly trade their personal information for free apps, promotional offers and social media incentives (Hull, 2014). In this case, Nunan and Di Domenico (2013,p.6) claim that "while privacy concerns have been raised over the use and creation of big data, these have been outpaced by

Electronic Markets - The International Journal on Networked Business
Editorial Office:
c/o Information Systems Institute
University of Leipzig, Germany
www.electronicmarkets.org

Grimmaische Str. 12
04109 Leipzig, Germany
Fax: +49 (341) 9733 612
editors@electronicmarkets.org

individuals' use of social networks". Although consumers increasingly share personal information in e-commerce sites or in social networks, it is expected for firms not to breach consumers' privacy because consumers disclose information expecting it to be confidential under 'terms of use'(Martin, 2015). Despite consumers' expectations for anonymous data to protect their privacy, the extant review on BDA identifies a new wave technologies and tools (e.g., facial recognition software) to de-anonymize and re-identify people in data economy. It raises serious concerns on the use of so called big public data (Boyd and Crawford, 2012). As a result, there is an urgent research call to protect privacy both technologically and legally in the era of biometric and genomic big data research (Kaplan, 2014).

In addition to privacy, big data creates serious security challenges as consumers are completely unaware of how their data are being used by whom and for what purposes. In this context, Vaidhyanathan and Bulock (2014) raised question on the validity of monitoring buying behaviour online and the extent to which consumers are knowledgeable and aware of this sort of monitoring. Firms within big data industry are often involved in creating an aggregated negative externality because they jointly contribute to the development of a large system of surveillance (Martin, 2015). Any such surveillance and potential revelations (e.g., Edward Snowden's revelation of PRISM program and its involvement with big corporations-Yahoo, Google, Facebook, Microsoft, Apple etc.) call attention to the security of private information (Bankston and Soltani, 2014; Schneier, 2013). In this context, Google has recently introduced "the right to be forgotten" policy in the European Union, which allows an individual to remove irrelevant personal data from its search results. The realm of big data-sharing agreements still remains informal, poorly structured, manually enforced, and linked to isolated transactions (George et al., 2014; Pantelis and Aija, 2013). Thus, to succeed in the

emerging big data environment, e-commerce firms need to be responsible to handle autonomy and informed consent and ensure privacy and security of data.

Another key challenge of the big data environment is to find the skills, such as technical, analytical, and governance skills as well as the networked relationships needed to operationalize big data (Davenport et al., 2012; Kiron et al., 2014a; Schroeck et al., 2012). However, it is not easy to find all these skills in one person. As argued by McAfee and Brynjolfsson (2012) and Kiron et al. (2014a), the enormous amount of big data needs to be captured, integrated, cleaned, and visualized; therefore, the technical and analytical skills of the data scientist (e.g., statistical, contextual, quantitative, predictive, and cognitive skills, and other related knowledge) are critical. In addition, these scientists should be conversant with business and governance issues, and should have the skills to communicate in the language of business. According to the sociomaterialism theory in IT (Orlikowski and Scott, 2008), organizational (i.e., BDA management), technological (i.e., IT infrastructure), and talent (e.g., analytics skill or knowledge) dimensions of analytics are so interwoven that it is difficult to measure their individual contribution in isolation (Orlikowski and Scott, 2008). Therefore, it is essential to develop big data analytics capability focusing on sophisticated technology, robust talent and analytics driven management culture.

Overall, the lack of organizational ability to articulate a solid and compelling business case is likely to be an overarching challenge for BDA in e-commerce. Both in academia and practice, researchers (e.g., Hayashi, 2014; Kiron et al., 2014b; Wamba et al., 2015) reported that a fascinating case for business opportunities with measureable benefits is the key challenge of big data. As most organizations are facing the dilemma of how to use big data, we advise that the first step is to spend time in creating a simple plan for how data, analytics, frontline tools, and people can come together to create business value (Biesdorf et al., 2013). Specifically, in

a similar spirit with Davenport (2013b), we suggest developing a research blueprint by recognizing the problem, reviewing past findings, identifying the variables and developing the model, collecting and analyzing the data and making decisions on actionable insights. Overall, there is wealth of possibilities to address exciting, non-trivial questions in e-commerce and we offer some illustrations in Table 4 focusing on specific research streams. However, the assessment of exciting, blue ocean research question in e-commerce should be based on fit, rigor, story, theory and economic and social significance.

8. Conclusion

Big data analytics (BDA) has emerged as the new frontier of innovation and competition in the wide spectrum of the e-commerce landscape due to the challenges and opportunities created by the information revolution. Big data analytics (BDA) increasingly provides value to e-commerce firms by using the dynamics of people, processes, and technologies to transform data into insights for robust decision making and solutions to business problems. This is a holistic process which deals with data, sources, skills, and systems in order to create a competitive advantage. Leading e-commerce firms such as Google, Amazon, and Facebook have already embraced BDA and experienced enormous growth. Through its systematic review and creation of taxonomy of the key aspects of BDA, this study presents a useful starting point for the application of BDA in emerging e-commerce research. The study presents an approach for encapsulating all the best practices that build and shape BDA capabilities. In addition, the study reflects that once BDA and its scope are well defined; distinctive characteristics and types of big data are well understood; and challenges are properly addressed, the BDA application will maximize business value through facilitating the pervasive usage and speedy delivery of insights across organizations.

References

Agarwal, R., Dhar, V., 2014. Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. Information Systems Research 25, 443-448.

Agarwal, R., Weill, P., 2012. The Benefits of Combining Data With Empathy. MIT Sloan Management Review 54, 35.

Allen, B., Bresnahan, J., Childers, L., Foster, I., Kandaswamy, G., Kettimuthu, R., Kordas, J., Link, M., Martin, S., Pickett, K., Tuecke, S., 2012. Software as a Service for Data Scientists. Communications of the ACM 55, 81-88.

Ann Keller, S., Koonin, S.E., Shipp, S., 2012. Big data and city living - what can it do for us? Significance 9, 4-7.

Bankston, K.S., Soltani, A., 2014. Tiny constables and the cost of surveillance: Making cents out of United States V. Jones. Yale Law Journal Online 123.

Barney, J., 1991. Firm resources and sustained competitive advantage. Journal of management 17, 99-120.

Barrett, M., Davidson, E., Prabhu, J., Vargo, S.L., 2015. Service Innovation in the Digital Age. MIS Quarterly 39, 135-154.

Barton, D., 2012. Making Advanced Analytics Work For You. Harvard business review 90, 78-83, 128.

Barton, D., Court, D., 2012. Making advanced analytics work for you. Harvard business review 90, 78.

Beath, C., Becerra-Fernandez, I., Ross, J., Short, J., 2012. Finding Value in the Information Explosion. MIT Sloan Management Review 53, 18-20.

Benedettini, O., Neely, A., 2012. Complexity in services: an interpretative framework, POMS 23rd Annual Conference.

Bennett, P., Giles, L., Halevy, A., Han, J., Hearst, M., Leskovec, J., 2013. Channeling the deluge: research challenges for big data and information systems, Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. ACM, pp. 2537-2538.

Beskow, L.M., Friedman, J.Y., Hardy, N.C., Lin, L., Weinfurt, K.P., 2010. Developing a simplified consent form for biobanking. PLoS One 5, e13302.

Beulke, D., 2011. Big Data Impacts Data Management: The 5 Vs of Big Data.

Bhattacherjee, A., 2001. Understanding information systems continuance: an expectation-confirmation model. MIS Quarterly 25, 351-370.

Bialobrzeski, A., Ried, J., Dabrock, P., 2012. Differentiating and evaluating common good and public good: Making implicit assumptions explicit in the contexts of consent and duty to participate. Public health genomics 15, 285-292.

Biesdorf, S., Court, D., Willmott, P., 2013. Big data: What's your plan?, McKinsey Quarterly.

Birnik, A., Bowman, C., 2007. Marketing mix standardization in multinational corporations: a review of the evidence. International Journal of Management Reviews 9, 303-324.

Boja, C., Pocovnicu, A., Batagan, L., 2012. Distributed Parallel Architecture for "Big Data". Informatica Economica 16, 116-127.

Bose, R., 2009. Advanced analytics: opportunities and challenges. Industrial Management & Data Systems 109, 155-172.

Bouhaddou, O., Bennett, J., Cromwell, T., Nixon, G., Teal, J., Davis, M., Smith, R., Fischetti, L., Parker, D., Gillen, Z., 2011. The Department of Veterans Affairs, Department of Defense, and Kaiser Permanente Nationwide Health Information Network Exchange in San Diego: Patient Selection, Consent, and Identity Matching, AMIA Annual Symposium Proceedings. American Medical Informatics Association, p. 135.

Boyd, D., Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information Communication and Society 15, 662-679.

Bragge, J., Sunikka, A., Kallio, H., 2012. An Exploratory Study on Customer Responses to Personalized Banner Messages in the Online Banking Context. JITTA : Journal of Information Technology Theory and Application 13, 5-18.

Braun, V., Clarke, V., 2006. Using Thematic Analysis in Psychology Qualitative Research in Psychology, 3, 77-101. Bristol: University of the West of England.

Brown, B., Chul, M., Manyika, J., 2011. Are you ready for the era of 'big data'? McKinsey Quarterly, 24-27+30-35.

Bughin, J., Chui, M., Manyika, J., 2010. Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. McKinsey Quarterly, 26-43.

Bughin, J., Livingston, J., Marwaha, S., 2011. Seizing the potential of 'big data'. McKinsey Quarterly, 103-109.

Chandrasekaran, S., Levin, R., Patel, H., Roberts, R., 2013. Winning with IT in consumer packaged goods: Seven trends transforming the role of the CIO. McKinsey & Company, pp. 1-8.

Chang, R.M., Kauffman, R.J., Kwon, Y., 2014. Understanding the paradigm shift to computational social science in the presence of big data. Decision Support Systems 63, 67-80.

Cherif, E., Grant, D., 2013. Analysis of e-business models in real estate. Electron Commer Res, 1-26.

Christian, H., 2013. Big data and the creative destruction of today's business models. AT Kerney, Available: http://www. atkearney. fr/strategic-it/ideas-insights/article/-/asset_publisher/LCcgOeS4t85g/content/big-data-and-the-creative-destruction-of-today-s-business-models/10192.

Columbus, L., 2014. Making Analytics Accountable: 56% Of Executives Expect Analytics to Contribute To 10% Or More Growth in 2014, Forbes.

Constantiou, I.D., Kallinikos, J., 2014. New games, new rules: big data and the changing context of strategy. Journal of Information Technology.

Davenport, T.H., 2006. Competing on analytics. harvard business review 84, 98-107.

Davenport, T.H., 2010. The New World of "Business Analytics. International Institute for Analytics, pp. 1-6.

Davenport, T.H., 2012. The Human Side of Big Data and High-Performance Analytics. International Institute for Analytics, pp. 1-13.

Davenport, T.H., 2013a. Analytics 3.0. Harvard Business Review 91, 64-72.

Davenport, T.H., 2013b. Keep Up with Your Quants. Harvard Business Review 91, 120-123.

Davenport, T.H., Barth, P., Bean, R., 2012. How 'Big Data' is Different. MIT Sloan Management Review 54, 43-46.

Davenport, T.H., Harris, J.G., 2007a. Competing on analytics: the new science of winning. Harvard Business School Press.

Davenport, T.H., Harris, J.G., 2007b. The dark side of customer analytics. Harvard Business Review 85, 37-+.

Davenport, T.H., Patil, D., 2012. Data scientist: the sexiest job of the 21st century. Harvard business review 90, 70-77.

De Swert, K., 2012. Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha. Center for Politics and Communication.

Delone, W.H., 2003. The DeLone and McLean model of information systems success: a ten-year update. Journal of Management Information Systems 19, 9-30.

DeLone, W.H., McLean, E.R., 1992. Information systems success: the quest for the dependent variable. Information Systems Research 3, 60-95.

Demirkan, H., Delen, D., 2012. Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. Decision Support Systems.

Devaraj, S., Fan, M., Kohli, R., 2002. Antecedents of B2C channel satisfaction and preference: validating e-commerce metrics. Information systems research 13, 316-333.

Dijcks, J.P., 2012. Oracle: Big Data for the Enterprise. Oracle.

Dimon, R., 2013. Understand: Turning Insights into Actions. Enterprise Performance Management Done Right: An Operating System for Your Organization, 57-77.

Dinev, T., Hart, P., 2006. An Extended Privacy Calculus Model for E-Commerce Transactions. Information Systems Research 17, 61-80.

emarketer, 2013. Ecommerce Sales Topped $1 Trillion for First Time in 2012

Ezzy, D., 2002. Qualitative analysis: practice and innovation Allen & Unwin. Crows Nest.

Ferguson, R.B., 2012. Risky Business: How Data Analytics and Behavioral Science Can Help. MIT Sloan Management Review 54, 1-5.

Fisher, D., DeLine, R., Czerwinski, M., Drucker, S., 2012. Interactions with Big Data Analytics. Interactions 19, 50.

Fosso Wamba, S., Akter, S., Coltman, T., W.T. Ngai, E., 2015a. Guest editorial: information technology-enabled supply chain management. Production Planning & Control 26, 933-944.

Fosso Wamba, S., Akter, S., Coltman, T., WT Ngai, E., 2015b. Guest editorial: information technology-enabled supply chain management. Production Planning & Control 26, 933-944.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., Gnanzou, D., 2015c. How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics 165, 234-246.

Fosso Wamba, S., Anand, A., Carter, L., 2013. A literature review of RFID-enabled healthcare applications and issues. International Journal of Information Management 33, 875-891.

Frost, R., Strauss, J., 2013. e-Marketing. Pearson Higher Ed.

Gefen, D., 2000. E-commerce: the role of familiarity and trust. Omega 28, 725-737.

Gefen, D., 2002. Customer loyalty in e-commerce. Journal of the Association for Information Systems (Volume 3, 2002) 27, 51.

Gehrke, J., 2012. Quo vadis, data privacy? Annals of the New York Academy of Sciences 1260, 45-54.

Gentile, B., 2012. Top 5 Myths About Big Data.

George, G., Haas, M.R., Pentland, A., 2014. Big data and management. Academy of Management Journal 57, 321-326.

Gobble, M.M., 2013. Big Data: The Next Big Thing in Innovation. Research Technology Management 56, 64-66.

Goes, P.B., 2014. Big Data and IS Research. MIS Quarterly 38, iii-viii.

Goff, J., McInerney, P., Soni, G., 2012. Need for speed: Algorithmic marketing and customer data overload, McKinsey Quarterly.

Griffin, J., Danson, F., 2012. Analytics and the Cloud - the Future is Here. Financial Executive 28, 97-98.

Griffin, R., 2012. Using Big Data to Combat Enterprise Fraud. Financial Executive 28, 44-47.

Havens, T.C., Bezdek, J.C., Leckie, C., Hall, L.O., Palaniswami, M., 2012. Fuzzy c-Means Algorithms for Very Large Data. Fuzzy Systems, IEEE Transactions on 20, 1130-1146.

Hayashi, A.M., 2014. Thriving in a Big Data World. MIT Sloan Management Review 55, 35-39.

Hayes, A.F., 2011. My macros and code for SPSS and SAS. Retrieved September 27, 2011.

Hayes, A.F., Krippendorff, K., 2007. Answering the call for a standard reliability measure for coding data. Communication methods and measures 1, 77-89.

Highfield, 2012. Talking of Many Things: Using Topical Networks to Study Discussions in Social Media.

Holbrook, M.B., 1999. Consumer value: a framework for analysis and research. Psychology Press.

Hsinchun, C., Chiang, R.H.L., Storey, V.C., 2012. BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. MIS Quarterly 36, 1165-1188.

Hull, G., 2014. Successful failure: what Foucault can teach us about privacy self-management in a world of Facebook and big data. Ethics and Information Technology, 1-13.

Huwe, T.K., 2012. Big data, big future. Computers in Libraries 32, 20-22.

IBM, 2012. What is big data?

IBM, 2013. Analytics: The real-world use of big data: How innovative retail organizations extract value from uncertain data.

IDC, 2013. Big Data in 2020, in: IDC (Ed.). IDC.

Ioannidis, J.P., 2013. Informed consent, big data, and the oxymoron of research that is not research. The American Journal of Bioethics 13, 40-42.

Jacobs, A., 2009. The Pathologies of Big Data. Association for Computing Machinery. Communications of the ACM 52, 36.

Jao, J., 2013. Why Big Data Is A Must In Ecommerce.

Johnson, B.D., 2012a. The Secret Life of Data. The Futurist 46, 20-23.

Johnson, J.E., 2012b. BIG DATA + BIG ANALYTICS = BIG OPPORTUNITY. Financial Executive 28, 50-53.

Kalakota, R., Whinston, A.B., 1997. Electronic commerce: a manager's guide. Addison-Wesley Professional.

Kang, K.-D., Son, S.H., Stankovic, J.A., 2003. Differentiated real-time data services for e-commerce applications. Electron Commer Res 3, 113-142.

Kaplan, B., 2014. Selling Health Data: De-Identification, Privacy, and Speech. Cambridge Quarterly of Healthcare Ethics, Forthcoming.

Kauffman, R.J., Srivastava, J., Vayghan, J., 2012. Business and data analytics: New innovations for the management of e-commerce. Electronic Commerce Research and Applications 11, 85-88.

Kim, G., Shin, B., Kwon, O., 2012. Investigating the Value of Sociomaterialism in Conceptualizing IT Capability of a Firm. Journal of Management Information Systems 29, 327-362.

Kiron, D., Prentice, P.K., Ferguson, R.B., 2012a. Innovating with Analytics. MIT Sloan Management Review 54.

Kiron, D., Prentice, P.K., Ferguson, R.B., 2014a. The analytics mandate. MIT Sloan management review 55, 1-25.

Kiron, D., Prentice, P.K., Ferguson, R.B., 2014b. Raising the Bar With Analytics. MIT Sloan Management Review 55, 29-33.

Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., Haydock, M., 2012b. Analytics: The Widening Divide. MIT Sloan Management Review 53, 1-22.

Koch, C., 2013. Compilation and synthesis in big data analytics, Big Data. Springer, pp. 6-6.

Kohli, A.K., Jaworski, B.J., 1990. Market orientation: the construct, research propositions, and managerial implications. The Journal of Marketing, 1-18.

Koirala, P., 2012. What is Big Data Analytics and its Application in E-Commerce? www.venturecity.com.

Kopp, M., 2013. Seizing the Big Data Opportunity, Ecommerce Times.

Koutsabasis, P., Stavrakis, M., Viorres, N., Darzentas, J.S., Spyrou, T., Darzentas, J., 2008. A descriptive reference framework for the personalisation of e-business applications. Electron Commer Res 8, 173-192.

Krippendorff, K., 2004. Reliability in content analysis. Human Communication Research 30, 411-433.

Krippendorff, K., 2007. Computing Krippendorff's alpha reliability. Departmental Papers (ASC), 43.

Kung, D.S., Gordon, L.C., Lin, F., Shayo, C., Dyck, H., 2013. Business Analytics: IT-based System with Dynamic Pricing Algorithm. Business Journal for Entrepreneurs 2013.

Lane, 2012. O Privacy, Where Art Thou?: Protecting Privacy and Confidentiality in an Era of Big Data Access.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N., 2011. Big Data, Analytics and the Path From Insights to Value. MIT Sloan Management Review 52, 21-32.

Leavitt, N., 2013. Bringing Big Analytics to the Masses. Computer 46, 20-23.

Leloup, B., Deveaux, L., 2001. Dynamic pricing on the internet: Theory and simulations. Electron Commer Res 1, 265-276.

Liebowitz, J., 2013. Big Data and Business Analytics. CRC Press.

Lim, E.P., Chen, H., Chen, G., 2013a. Business intelligence and analytics: Research directions. ACM Transactions on Management Information Systems 3.

Lim, M.K., Bahr, W., Leung, S.C.H., 2013b. RFID in the warehouse: A literature analysis (1995-2010) of its applications, benefits, challenges and future trends. International Journal of Production Economics.

Loebbecke, C., Picot, A., 2015. Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. The Journal of Strategic Information Systems 24, 149-157.

Loveman, G., 2003. Diamonds in the data mine. Harvard business review 81, 109-113.

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.

Martin, K.E., 2015. Ethical Issues in the Big Data Industry. MIS Quarterly Executive, Forthcoming.

McAfee, A., Brynjolfsson, E., 2012. Big data: the management revolution. Harvard business review 1.

Mehra, G., 2013. 6 Uses of Big Data for Online Retailers, Practical Ecommerce.

Melville, N., Kraemer, K., Gurbaxani, V., 2004. Review: Information technology and organizational performance: An integrative model of IT business value. MIS quarterly 28, 283-322.

Miller, G., 2013. 6 Ways To Use "Big Data" To Increase Operating Margins By 60%.

Minelli, M., Chambers, M., Dhiraj, A., 2013. Business Analytics. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, 99-125.

Mithas, S., Lee, M.R., Earley, S., Murugesan, S., Djavanshir, R., 2013. Leveraging Big Data and Business Analytics. IT Professional 15, 18-20.

Narver, J.C., Slater, S.F., 1990. The effect of a market orientation on business profitability. The Journal of Marketing, 20-35.

Nelson, R.R., Todd, P.A., Wixom, B.H., 2005. Antecedents of information and system quality: an empirical examination within the context of data warehousing. Journal of Management Information Systems 21, 199-235.

Ngai, E.W.T., Moon, K.K.L., Riggins, F.J., Yi, C.Y., 2008. RFID research: An academic literature review (1995-2005) and future research directions. International Journal of Production Economics 112, 510-520.

Ngai, E.W.T., Wat, F.K.T., 2002. A literature review and classification of electronic commerce research. Information & Management 39, 415-429.

Ngai, E.W.T., Xiu, L., Chau, D.C.K., 2009. Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications 36, 2592-2602.

Nunan, D., Di Domenico, M., 2013. Market research and the ethics of big data. International Journal of Market Research 55, 505-520.

Ohata, M., Kumar, A., 2012. Big Data: A Boon to Business Intelligence. Financial Executive 28, 63-64.

Orlikowski, W.J., 2007. Sociomaterial practices: Exploring technology at work. Organization studies 28, 1435-1448.

Orlikowski, W.J., Scott, S.V., 2008. 10 Sociomateriality: Challenging the Separation of Technology, Work and Organization. The academy of management annals 2, 433-474.

Pantelis, K., Aija, L., 2013. Understanding the value of (big) data, 2013 IEEE International Conference on Big Data.

Piatetsky, G., 2014. Big Data market to reach $50 Billion by 2018. Wikibon.

Porter, M.E., Millar, V.E., 1985. How information gives you competitive advantage. Harvard Business Review, Reprint Service.

Rajpurohit, A., 2013. Big data for business managers—Bridging the gap between potential and value, Big Data, 2013 IEEE International Conference on. IEEE, pp. 29-31.

Ramaswamy, S., 2013. What the Companies Winning at Big Data Do Differently. Bloomberg, June: http://www. bloomberg. com/news/2013-06-25/what-the-companies-winning-at-big-data-do-differently. html.

Riggins, F.J., 1999. A Framework for Identifying Web-Based Electronic Commerce Opportunities. Journal of Organizational Computing and Electronic Commerce 9, 297-310.

Rouse, M., 2011. big data (Big Data).

Russom, P., 2011. The Three Vs of Big Data Analytics. TDWI.

Schneier, B., 2013. The US government has betrayed the internet. We need to take it back. The Guardian. Sept 5, 2013.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P.P., 2012. Analytics: The real-world use of big data. IBM Institute for Business Value, NY, USA.

Shah, S., Horne, A., Capellá, J., 2012. Good data won't guarantee good decisions. harvard business review 90, 23-25.

Shanks, G., Sharma, R., Seddon, P., Reynolds, P., 2010. The impact of strategy and maturity on business analytics and firm performance: A review and research agenda.

Sharma, R., Mithas, S., Kankanhalli, A., 2014. Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. Eur J Inf Syst 23, 433-441.

Smith, M., Szongott, C., Henne, B., Von Voigt, G., 2012. Big data privacy issues in public social media. IEEE International Conference on Digital Ecosystems and Technologies.

Smith, R., Shao, J., 2007. Privacy and e-commerce: a consumer-centric perspective. Electron Commer Res 7, 89-116.

Spiggle, S., 1994. Analysis and interpretation of qualitative data in consumer research. Journal of consumer research, 491-503.

Strawn, G.O., 2012. Scientific Research: How Many Paradigms? EDUCAUSE Review 47, 26.

Tankard, C., 2012. Big data security. Network Security 2012, 5-8.

The Economist, 2011. Building with big data: The data revolution is changing the landscape of business.

Vaidhyanathan, S., Bulock, C., 2014. Knowledge and Dignity in the Era of "Big Data". The Serials Librarian 66, 49-64.

Vaithianathan, S., 2010. A review of e-commerce literature on India and research agenda for the future. Electron Commer Res 10, 83-97.

Viaene, S., Van den Bunder, A., 2011. The Secrets to Managing Business Analytics Projects. MIT Sloan Management Review 53, 65-69.

Wagner, E., 2012. Realities learning professionals need to know about analytics. T and D 66, 54-58.

Waller, M.A., Fawcett, S.E., 2013. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. Journal of Business Logistics 34, 77-84.

Wamba, S.F., Akter, S., Edwards, A., Chopin, G., Gnanzou, D., 2015. How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics 165, 234-246.

White, M., 2012a. Digital workplaces Vision and reality. Business Information Review 29, 205-214.

White, M., 2012b. Digital workplaces: Vision and reality. Business Information Review 29, 205-214.

Williamson, O.E., 1979. Transaction-cost economics: the governance of contractual relations. Journal of law and economics, 233-261.

Williamson, O.E., 1981. The economics of organization: The transaction cost approach. American journal of sociology, 548-577.

Williamson, O.E., 2000. The new institutional economics: taking stock, looking ahead. Journal of economic literature, 595-613.

Wixom, B.H., Todd, P.A., 2005. A theoretical integration of user satisfaction and technology acceptance. Information Systems Research 16, 85-102.

Wixom, B.H., Yen, B., Relich, M., 2013. Maximizing value from business analytics. MIS Quarterly Executive 12, 111-123.

Zeng, D., Lusch, R., 2013. Big Data Analytics: Perspective Shifting from Transactions to Ecosystems. IEEE Intelligent Systems 28, 2-5.

Zhao, D., 2013. Frontiers of Big Data Business Analytics: Patterns and Cases in Online Marketing. Big Data and Business Analytics, 43.

Appendix 1: Definitional aspects of big data analytics (BDA) in e-commerce

Study	Potential research areas	Definition	Purpose
Davenport (2006)	E-commerce functions: production and operations (e.g., supply chain flows), marketing (e.g., promotion), human resources (e.g., employee performance), finance (e.g., controlling fraud), and research and development (R&D).	BDA refers to 'quantitative fact-based' analysis aiding decision making.	Analytics to gain competitive advantage.
Davenport and Harris (2007b)	E-commerce functions: finance (e.g., merger & acquisition), manufacturing, R&D, human resources (e.g., hire, retain, and promote the best people), marketing (e.g., identifying profitable and loyal customers), and supply chain (e.g., lowest possible inventory without compromising stockouts).	BDA refers to the use of data, statistical and quantitative analysis through explanatory and predictive models for facilitating fact-based management decisions and actions.	Analytics to help build distinctive capabilities in an intensely competitive business environment.
Bose (2009)	E-commerce functions including: marketing (direct marketing, customer segmentation, pricing), supply chain (choice of channel partners), and finance (customer profitability analysis).	BDA refers to a group of tools used in combination to gain insights in order to direct, optimize, and automate decision making for achieving organizational goals.	Assisting business managers to effectively understand the drivers behind advanced analytics implementation.
Shanks et al. (2010)	Functional areas including: production and operations (sales forecasts, production plans, order deliveries) and marketing (customer attrition, customer profitability, response rate of marketing campaigns, differential pricing, etc.).	BDA refers to data interpretation to generate insights that improve decision making, and optimize business processes.	Technology and capabilities using analytics lead to value-creating actions to improve firm performance and competitive advantage.
Manyika et al. (2011)	All e-commerce functions including: marketing, operations, and the supply chain.	BDA creates value by creating transparency, discovering needs, exposing variability, and improving performance.	Potential value of big data for organizations and the economy, outlining the ways to capture that value.
LaValle et al. (2011)	All e-commerce functions including: marketing (loyalty/retention/defection), finance (e.g., budgeting, revenue growth, cost efficiency), human resources management (workforce planning/allocation), operations and production.	BDA is about insights—descriptive, predictive, and prescriptive—to deliver actions that are closely linked to business strategy and to organizational processes that take place at the right time.	Insightful evidence to help organizations understand the opportunity provided by information and advanced analytics.

Electronic Markets - The International Journal on Networked Business   Grimmaische Str. 12
Editorial Office: 04109 Leipzig, Germany
c/o Information Systems Institute   Fax: +49 (341) 9733 612
www.electronicmarkets.org   University of Leipzig, Germany   editors@electronicmarkets.org

Agarwal and Weill (2012)	Data in conjunction with emotion significantly strengthen the firm's analytical ability for marketing (e.g., customer empowerment), human resources (e.g., employee empowerment), and supply chain (channel partner empowerment and relationship).	BDA improves business processes, emotional connections, and evidence based on empathetic use of data.	Integration of human emotion in data analytics in addition to process automation and optimization.
Davenport (2012)	BDA in e-commerce can significantly impact on product/service development (new features on online social site), marketing (ad-serving, pricing, trends on customer sentiments), and process improvement (price optimization time, service contracts, and maintenance optimization).	BDA focuses on discovering products, features, and value-adding services.	Identifying the quality and skill sets required for people dealing with big data in order to deliver high performance analytics (HPA).
Davenport et al. (2012)	BDA in e-commerce aids marketing (e.g., customization, on-demand pricing), R&D (e.g., discovery of a new drug), and production and operations.	BDA is used to seek continuously changing patterns, events, and opportunities to generate discovery and agility.	Revealing the distinguishing facts between big data analytics and traditional analytics, and the ways to capitalize on insights from big data.
Dijcks (2012)	BDA in e-commerce influences manufacturing (telemetry reveals usage patterns, failure rates, and other opportunities for product improvement), marketing (smartphones and other GPS devices offer advertisers proximity information, effective micro customer segmentation and targeted marketing campaigns), and supply chain (efficiency).	BDA supports deeper analysis on a wider variety of data types, delivering faster response times driven by changes in behavior and automating decisions based on analytical models.	The importance of and infrastructure required for business analytics in the big data environment.
Ferguson (2012)	BDA in e-commerce aids operational risk assessment involving finance (internal fraud, external fraud, damage to physical assets), marketing (customer/client intentions and behaviors, products and business practices), human resources (employment practices), and information technology (system failure).	BDA can be said to be the multidimensional behavioral analysis used to discover failure modes in advance of a disaster.	Explains how analytics helps organizations better understand and mitigate everything from risky business practices to the big undetectable risks.
(Kiron et al., 2012a, 2014a)	BDA in e-commerce influences decisions about marketing (customer care, customer experience, pricing strategy, sales, customer service), finance (productivity, mergers and acquisitions), and operations.	BDA refers to the application of data and business insights developed through applied analytical disciplines to drive fact-based planning, decisions, execution, management, measurement, and learning.	Explanations of the practical concerns that organizations are facing regarding the use of analytics.

Appendix 2: Nature of big data used in business analytics

Nature	Description	Example (Study)
Voluminous	Large volume of data that either consume a huge amount of storage or consist of a large number of records (Davenport et al., 2012; Russom, 2011).	• Amazon introduced search-inside-the-books option consisting of 120,000 books (Davenport, 2006). • Dell initiated the development of a database that includes 1.5 million records related to sales and advertisements (Davenport, 2006). • Netflix analyzes customer choice and customer feedback from over one billion reviews (Davenport and Harris, 2007a). • Match.com has billions of data points to analyze generated from the past 17 years (Kiron et al., 2012b). • On Facebook, 30 billion pieces of content are shared every month (Manyika et al., 2011). • Tesco generates more than 1.5 billion new items of data every month (Manyika et al., 2011). • Wal-Mart's data warehouse includes some 2.5 petabytes of information (Manyika et al., 2011).
Variety	Data generated from a greater variety of sources and formats, and that contain multidimensional data fields (Davenport et al., 2012; Russom, 2011).	• The DDB Matrix database includes all print, radio, network TV, and cable ads of Dell, the computer manufacturer, as well as regional sales data (Davenport, 2006). • Credit card companies use website click-stream data and other data formats from call center operations to customize offers (Davenport and Patil, 2012). • A retailer's analytic model may include customer profiles and purchase history, regional and seasonal buying patterns, optimizing of supply chain operations, and unstructured data from social media to customize predictions by product, store concept, and promotional campaigns, etc. (Biesdorf et al., 2013). • Cross-selling uses all possible data that can be identified and stored about a customer, including the customer's demographics, purchase history, preferences, real-time locations, and other facts to increase the average purchase size (Manyika et al., 2011). • Retailers use sentiment analysis to assess the real-time response to marketing campaigns and to make

		adjustments as needed. The evolving field of social media analytics plays a key role as consumers are relying increasingly on peer sentiment and recommendations (Manyika et al., 2011).
		• Research identified 10 distinct methods in which to interact with consumers on social media throughout the product decision-making process, which range from passive techniques (monitoring blogs and social networks for references to brands) to direct engagement in the form of targeted marketing, new-product introductions, or consumer outreach during public relations (Chandrasekaran et al., 2013).
Velocity	Frequency of data generation and/or frequency of data delivery (Russom, 2011).	• Amazon manages a constant flow of new products, suppliers, customers, and promotions without compromising its promised delivery dates (Davenport, 2006).
		• Consumer sentiment analysis through social media analytics requires real-time monitoring of the environment. The frequent flow of new data makes the decision role frequently obsolete. Trends in customer sentiments about products, brands, and companies are of high velocity with a larger volume (Davenport, 2012).
		• At Twitter, even at 140 characters per tweet, the data volumes are estimated at eight terabytes per day due only to the high velocity/frequency (Dijcks, 2012).
		• Retailers can now track the individual customer's data, including click-stream data from the web, and can leverage based on their behavioral analysis. Moreover, retailers are now capable of updating such increasingly granular data in near real time to track changes in customer behavior (Manyika et al., 2011).
		• eBay Inc. conducts thousands of experiments with different aspects of its website to determine optimal layout and other features ranging from navigation to the size of its photos (Manyika et al., 2011).

Veracity	Generating authenticated and relevant data with the capability of screening out bad data (Beulke, 2011).	• eBay Inc. faced enormous data replication problems, with between 20-fold and 50-fold versions of the same data scattered throughout its various data marts. Later, eBay Inc. developed an internal website (a data hub) that enabled managers to filter data replication (Davenport et al., 2012). • Using data fusion, an organization can combine multiple less reliable sources to create a more accurate and useful data point, such as social comments affixed to geospatial location information (Schroeck et al., 2012). • 'Black swans' (the disproportionate quality of high-impact, hard-to-predict, rare events that generally people do not expect to happen, which are also said to be extreme outliers) are outside the realm of detection—by quantitative, traditional analytics. Montage Analytics has developed a tool that can predict areas vulnerable to 'black swans' within organizations, and other types of risk involving the impact of human behavior and motivations (Ferguson, 2012).

Appendix 3: Types of big data

Types	Descriptions	Applications by e-Businesses/Firms
Transaction or business activity data	Structured data from retail transactions, customer profiles, distribution frequency and volume, product consumption and service usage, nature and frequency of customer complaints	• United Parcel Service (UPS) examines usage patterns and complaints to predict customer defection (Davenport, 2006). • Wal-Mart persuades its suppliers to monitor product movement by store to help plan promotions, store layout, and reduce stockouts (Davenport, 2006). • Amazon engages a type of predictive modeling technique called collaborative filtering, using customer data to generate 'you might also want' prompts for each product bought or visited. Amazon revealed at one point that 30% of sales were generated through its recommendation engine (Manyika et al., 2011). • Harrah's, the US hotels and casinos group, compiles detailed holistic customer profiles and uses them to customize marketing in a way that has increased customer loyalty (Manyika et al., 2011). • Progressive Insurance and Capital One are conducting experiments on a regular basis to segment their customers systematically and effectively and to personalize product offers (Manyika et al., 2011). • Wal-Mart developed Retail Link, a tool that presents its suppliers with a view of the demand in its stores so suppliers know when stores should be restocked rather than waiting for an order from Wal-Mart stores (Manyika et al., 2011).
Click-stream data	Click-stream data from the web, social media content, online advertisements (tweets, blogs, Facebook wall postings,	• Every day, Google alone processes about 24 petabytes (or 24,000 terabytes) of data (Davenport, 2012). • Netflix Inc. analyzes web data of over one billion reviews of movies that were liked, loved, hated, etc. to recommend movies that optimize customers' tastes and inventory conditions

		etc.)	(Davenport and Harris, 2007b).
			• Credit card companies use a 'ready-to-market' database of website and call center activities to make personalized offers in milliseconds and to optimize offers by tracking responses (Davenport, 2012).
			• Merchandise buyers of a retail firm receive a red flag alert when competitors' internet sites price products below their retail firm's level (Biesdorf et al., 2013).
			• Retailers use sentiment analysis to assess the real-time response to marketing campaigns and make adjustments as needed. The evolving field of social media analytics plays a key role as consumers are relying increasingly on peer sentiment and recommendations (Manyika et al., 2011).
			• eBay Inc. conducts thousands of experiments with different aspects of its website to determine optimal layout and other features ranging from navigation to the size of its photos (Manyika et al., 2011).
Video data	Video data from retail and other settings		• Less than 25% of business and IT professionals with active big data efforts reported that they have the required capabilities to analyze extremely unstructured data such as voice and video (Schroeck et al., 2012).
			• Some retailers utilize sophisticated image-analysis software linked to their video-surveillance cameras to track in-store traffic patterns and consumer behavior (Manyika et al., 2011).
			• A consumer goods manufacturer equipped its merchandisers with tablet apps that enable the use of pictures and data to track shelf-space allocation compared to that of competitors on a daily basis and also to detect retailers' compliance with promotion agreements (Chandrasekaran et al., 2013).
			• P&G implemented the use of design tools to create realistic virtual prototypes that lead to

		time savings in design iterations. In addition, the virtual-reality techniques allowed the simulation of new products placed on shelves in order to test design effects internally and with consumers (Chandrasekaran et al., 2013).
Voice data	Voice data from phone calls, call centers, customer service	• Credit card companies use and track call center activities to make personalized offers in milliseconds and to optimize offers by tracking responses (Davenport, 2012). • Respondents from e-business enterprises (more than half of those with active big data efforts) reported that they are using advanced capabilities to analyze text in its natural state, such as the transcripts of call center conversations. These analytics have the ability to interpret and understand the nuances of language, such as sentiment, slang and intentions (Schroeck et al., 2012).