Exploring how consumer goods companies innovate in the digital age: the role of big data analytics companies

Article

Accepted Version

Mariani, M. ORCID: https://orcid.org/0000-0002-7916-2576 and Fosso Wamba, S. (2020) Exploring how consumer goods companies innovate in the digital age: the role of big data analytics companies. Journal of Business Research, 121. pp. 338-352. ISSN 0148-2963 doi: 10.1016/j.jbusres.2020.09.012 Available at https://centaur.reading.ac.uk/93071/

To link to this article DOI: http://dx.doi.org/10.1016/j.jbusres.2020.09.012

Publisher: Elsevier

Exploring how consumer goods companies innovate in the digital age: The role of big data analytics companies

Abstract The advent and development of digital technologies have brought about a proliferation of online consumer reviews (OCRs), i.e., real-time customers' evaluations of products, services, and brands. Increasingly, e-commerce platforms are using them to gain insights from customer feedback. Meanwhile, a new generation of big data analytics (BDA) companies are crowdsourcing large volumes of OCRs by means of controlled ad hoc online experiments and advanced machine learning (ML) techniques to forecast demand and determine the market potential for new products in several industries. We illustrate how this process is taking place for consumer goods companies by exploring the case of UK digital BDA company, SoundOut. Based on an in-depth qualitative analysis, we develop the consumer goods company innovation (CGCI) conceptual framework, which illustrates how digital BDA firms help consumer goods companies to test new products before they are launched on the market, and innovate. Theoretical and managerial implications are discussed.

Keywords: Big data analytics; forecasting; innovation; online review crowdsourcing; consumer goods companies; digital data.

Acknowledgements:

We would like to express our gratitude to all the interviewees (especially to Mr. David Courtier Dutton) for their availability, and would like to thank the Associate Editors and the anonymous reviewers for their helpful and constructive comments. We are grateful to two colleagues for their advice on how to improve our paper and to David Inglis and Kathryn Pilgrem for their assistance with professional copyediting and design. Last, we would like to thank SoundOut for generously providing copyright clearance for Figures 1-3 included in this work.

1. Introduction

Individuals and organizations are currently navigating in more dynamic, uncertain, and complex economic environments where technological, socio-political, and environmental changes recurrently challenge any competitive advantage created, making it transient and temporary (D'Aveni, 2010). Mounting uncertainty leads firms to devise strategies, tactics, and operational tools to recreate, at a fast pace, a competitive advantage once it has been eroded (D'Aveni, 1994, 1995). Increasingly, strategies, tactics, and tools that enable firms to recreate their competitive advantage rely on big data that can be conducive to the generation of operational and strategic value (Davenport et al., 2012; Dean, 2014; Fosso Wamba et al., 2015; Ghasemaghaei & Calic, 2020; Mariani & Borghi, 2019; Mikalef et al, 2019a). Accordingly, an increasing number of firms have been competing on, and leveraging, big data analytics (BDA) (Davenport, 2006) in order to create, deliver, and capture business value, while deriving IT infrastructure, operational, managerial, strategic, and organizational benefits (Wang et al., 2018) that can ultimately be translated into competitive advantage and better performance (Akter et al., 2016; Fosso Wamba et al, 2015, 2017; Mariani et al., 2018).

In other words, today's hypercompetitive business arena is a complex series of concatenated business ecosystems (Kapoor & Lee, 2013; Tsujimoto et al., 2018), where multiple economic actors compete, cooperate or even coopete (Brandenburger & Nalebuff, 1996) to control and analyze large volumes of data. In this regard, BDA appears as a source of competitive yet temporary advantage in the global and digital economy.

So far, most management scholars have focused on value creation in terms of improving decision-making through automated algorithms, customer micro-segmentation, and pricing decisions. Accordingly, the extant literature has focused on several areas, including statistical algorithms, marketing analytics, and customer analytics (Batistič & van der Laken, 2019). However, very few of them (Kakatkar et al., 2020) have empirically addressed the issue of how to leverage BDA for product, process, and business model innovation, and more specifically, the role and features of innovation analytics. This area is particularly important, especially for consumer goods companies (CGCs) operating in fast-changing and dynamic markets such as fashion and entertainment, as these firms must continuously innovate to maintain their competitive edge. However, compared to (large) financial companies that have endowed themselves with internal BDA departments, functions, and taskforces, as well as launched fin-tech initiatives (Bean, 2018), CGCs have been slower to integrate BDA into their operations so as to experiment and test new products and stand among the leading adopters of BDA (Dresner Advisory Services, 2018).

As such, and despite scattered anecdotal evidence in practitioners' outlets (Beath et al., 2012; Bughin et al., 2011; Davenport, 2017; LaValle et al., 2011) and the emerging literature on big data (Merendino et al, 2018, Mikalef et al, 2019a), it appears that there is a knowledge gap revolving around how CIOs and innovation managers in CGCs are using BDA to innovate their products, services, and business models. Moreover, it is not clear if they are using BDA capabilities (BDAC) (Ferraris et al., 2019; Mikalef et al., 2020) in-house or if they outsource them for product innovation. Bridging this knowledge gap required that we conduct a qualitative case study of the UK-based company SoundOut, a leader in the provision of BDA to CGCs for innovation purposes. This consisted of: (1) describing the role of digital BDA companies in helping CGCs innovate their products and business models (the "*what*" question); (2) assessing *how* BDA companies generate predictive analytics that helps CGCs forecast demand for new products in rapidly changing markets (the "*how*" question); and (3) finding the reasons *why* some CGCs outsource predictive analytics for both product testing and innovation purposes.

Therefore, the main objective of this study is to bridge the aforementioned knowledge gap by analyzing how CGCs are increasingly adopting BDA and BDAC for new product testing and innovation purposes. As such, the study aims to examine the following research question: "As CGCs increasingly rely on professional BDA companies to capture business insights that are critical to innovate their products and services, *how* have they been helped and what types of BDA and BDAC contributed to help them?" By means of a revelatory case study, this question is being addressed, and the findings, along with the development of a novel conceptual framework, namely the consumer goods company innovation (CGCI) BDA framework, enabled us to enrich the extant BDA research. The CGCI BDA framework is particularly useful to understand how online consumer reviews (OCRs) are crowdsourced from consumers and leveraged using machine learning (ML) tools to predict consumers' demand for a new market product/service.

To make the aforementioned contributions, the paper is organized as follows. In section 2 we review the relevant literature: (i) studies about BDA as a source of business intelligence in the digital age; and (ii) online reviews and user generated content (UGC) as a source of business intelligence for CGCs in the digital age. Section 3 illustrates the methodology adopted. Section 4 presents the findings. The fifth section discusses the findings and presents theoretical and managerial implications. The last section portrays the limitations of the study while proposing an agenda for future research.

2. Theoretical background

2.1 Big data analytics as a source of business intelligence in the digital age

Research on data in social sciences has a relatively recent tradition, and "big data" has been increasingly studied over the last decade. Historically, the "big data" notion appeared in the late nineties in the computer science literature in relation to scientific visualization (Cox & Ellsworth, 1997). However, its first definition with visibility in the business domain was given in 2001 by Doug Laney, who identified three major characteristics of big data as the 3Vs

(Laney, 2001): Volume (size of data, today in the order of zettabytes), Velocity (rapidity of data generation, modification, and transfer), and Variety (data can assume different formats/structures). Later, the definitional model was perfected by introducing the Vs of Value (the process of extracting valuable knowledge from data by means of BDA) and Veracity (the governance of data in relation to their reliability), thus leading to the formulation of a 5Vs framework (Fosso Wamba et al, 2015).

There are multiple sources of big data that could originate from information searches (e.g., through search engines and meta-search engines), financial transactions (e.g., via e-commerce websites transactions and retail scanners), non-financial transactions (e.g., through services provided via e-government solutions/platforms), information diffusion (e.g., via information spread through company websites and apps), and social interactions (e.g., through UGC in the form of online review platforms, social networking sites, and blogs) (Blazquez & Domenech, 2018). Other sources of data are not created intentionally by internet users and include internet usage data (e.g., generated through web cookies), location data (created mostly through smartphones in the form of GSM, GPS, call detail records, and Bluetooth signals), and personal data (retrieved by websites through the type of searches or purchases).

Interestingly, big data allows firms and entrepreneurs to generate BDA, which is defined 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" (Fosso Wamba et al., 2020). BDA has been found to enhance business intelligence, thus allowing companies to improve customer satisfaction, optimize inventories, manage supply chain risks effectively, streamline operations, and identify target customer segments (Davenport, 2014, 2017). Such actions are conducive to superior levels of organizational performance and agility, as well as competitive advantage (Nam et al., 2019; Williams, 2016).

Nonetheless, BDA is not sufficient per se, as a firm needs to possess BDAC to make a difference in terms of performance (Sena et al., 2019) while facing institutional pressures (Dubey et al., 2019). Drawing on Gupta and George (2016), Mikalef et al (2020, p. 2) define BDAC as “the ability of a firm to effectively deploy technology and talent to capture, store, and analyze data, toward the generation of insight”. BDAC can allow firms to improve not only their performance (Ferraris et al, 2019; Mikalef et al, 2019b, 2020; Rialti, Zollo, Ferraris, & Alon, 2019), but also their supply chains (Srinivasan & Swink, 2018; Fosso Wamba et al., 2020)

CGCs have embraced and adopted BDA, with different speeds and scope, with a view to listening to their customers in real time and tailoring location-based offers and marketing communications. However, applications of BDA for CGCs are mostly related to location-based marketing communications, customer micro-segmentation, consumer attitudes, pricing, loyalty, optimal selling sequence, and distribution optimization (Chiang & Yang, 2018; Jun et al., 2014; Mariani, 2019; Shen et al., 2017). Accordingly, they have been mostly developing *marketing analytics*. For instance, Chiang and Yang (2018) use big data to examine 44,000 point-of-sale transaction records for 26,000 customers of a Taiwanese retail store to gain insights into how consumer personality traits are influenced by the country-of-origin features of beer brands, to predict potential customer lifetime value. They find that consumers tend to buy brands whose country-of-origin personality matches consumers’ personality traits. These findings mean it is possible to identify the most profitable customers for the brand analyzed. Shen et al (2017) developed a Bayesian information inventory model in order to help retailers to identify the optimal selling sequence of green and non-green products while improving on profit and environmental costs. Scholars interested in empowering CGCs to make better decisions have increasingly identified a suitable source of BDA in UGC and, especially, in

OCRs. In section 2.2, we review relevant literature focusing on BDA from OCRs to help improve decision-making processes and value creation for CGCs.

2.2 Online consumer reviews and user generated content as a source of business intelligence for CGCs in the digital age

Over the last 15 years, the proliferation of OCRs and their relevance in today's digital economy have attracted the attention of researchers from a range of disciplines, including data science, statistics, information and computer science, psychology, marketing, management science, etc. Researchers in data science and information and computer science have focused mostly on describing and analyzing the properties and features of OCRs (both quantitative and qualitative), and making sense of them by leveraging techniques such as data mining, ML, sentiment analysis, and more traditional statistical techniques (Chong et al., 2016; Clemons et al., 2006; Duan et al., 2008; Jones et al, 2004; Mariani, Di Fatta and Di Felice, 2019; Mariani & Predvoditeleva, 2019; Mariani & Borghi, 2020; Mariani et al., 2020; Mudambi & Shuff, 2010). Relevant OCR features examined are either quantitative or qualitative. The former set includes review valence (i.e., ratings), volume (i.e., number of reviews), variance (i.e., the variation of the ratings), and helpfulness (i.e., count of helpful votes). The latter encompasses review quality, identity disclosure, readability, informativeness, style, and sentiment.

Marketing and management scholars have focused not only on the mechanisms through which consumers produce and adopt OCRs (Hennig-Thurau et al., 2004), but also on the impact of OCRs on consumer decision-making (e.g., satisfaction, trust, attitude, booking intention, customer experience) and business performance (e.g., revenues and profits). Several of the studies related to the antecedents of OCRs have looked at the features that can enhance trustworthiness and credibility. Such features include reviewer and review quality, reviewer identity disclosure, helpful votes received in the past, and the number of reviews already written (e.g., Forman & Ghose, 2008;). Other studies have instead focused on understanding which factors (i.e., valence, volume, text sentiment, readability, informativeness, sequence of negative helpful reviews, helpful ratio, submission device), either individually or collectively, affect positively firm performance in terms of sales and revenues, and consumer attitudes and satisfaction (e.g., Basuroy et al., 2003; Chen et al, 2012b; Chevalier & Mayzlin, 2006; Chintagunta et al, 2010; Duan et al, 2008; Jun et al, 2014; Kim et al., 2017; Mariani, Borghi and Kazakov, 2019; Mariani, Borghi and Gretzel, 2019; Mariani & Borghi, 2020). Regarding the impact on sales, several scholars have found that the higher the valence (i.e., ratings) and volume of OCRs, the higher the product sales (e.g., Chevalier & Mayzlin, 2006; Chintagunta et al, 2010; Clemons et al, 2006; Duan et al, 2008) and firms' efficiency (Mariani & Visani, 2019). Others have found a positive relationship between qualitative features of OCRs—review quality and reviewer's characteristics (Lee & Shin, 2014)—and revenues or sentiment and sales (e.g., Chong et al, 2016; Hu et al., 2014). In general, the effect of both quantitative and qualitative features of OCRs on sales is also dependent on contextual factors such as product category, product characteristics, consumer characteristics, industry characteristics, platform characteristics, and firm strategic actions (Cui et al., 2012; You et al., 2015).

Currently, the way digital data streams from OCRs are retrieved, collected, processed, analyzed, reported, and visualized (George et al., 2016; Liu et al. 2019) differs across companies and organizations. In general, companies retrieving, processing, analyzing, reporting, and visualizing big data from OCRs have a great opportunity to accumulate business intelligence for their customers and markets, while being in a better position to optimize the activities of online review platforms. In particular, there are a few major knowledge gaps that can be bridged by effectively using BDA sourced from OCRs:

1) *The customer satisfaction knowledge gap*. Currently, major e-commerce platforms and online retailers can capture almost in real time the level of satisfaction/dissatisfaction

with a product or a service based on the analytics-based quantitative features of OCRs such as valence and sentiment. For instance, Booking.com and its almost 30 million partner accommodation providers (hotels and other lodging companies) can measure almost in real time the customer's satisfaction with hospitality service attributes such as cleanliness, comfort, facilities, staff, value for money, Wi-Fi, and location;

2) *The customer journey and behavior knowledge gap*. In most of the industries, marketing managers are increasingly interested in tracking the online customer journey and behavior by triangulating multiple data sources stemming not only from web clickstreams, transaction records, and voice recordings from call centers, but also from the richness of the content and features of OCRs. This triangulation allows them to capture the multi-channel journey of customers. For instance, the e-commerce platform Amazon enables reviewers to vote on the helpfulness of a review, thus contributing to customer journey improvement and facilitating navigation in a complex and expanding number of product reviews. This is the reason why the platform has developed algorithms allowing web users to filter the reviews based on their level of helpfulness. Predictive and prescriptive analytics company, 8451 explores OCR data to understand customer behavior with the purpose of strengthening customers' loyalty to brands while designing and carrying out targeted communications driving incremental purchases;

3) *The pricing knowledge gap*. While pricing displays a long tradition in the way analytics has been applied to manage revenues (see, for instance, the airline sector), OCRs can provide additional quantitative and qualitative information for optimizing prices. For instance, companies that enjoy higher ratings in OCRs can be more flexible in setting their prices. Increasingly, OCR data analytics companies such as SoundOut help business customers and merchandisers to ensure a better combination of high appeal and optimum price;

4) *The competitive and market intelligence knowledge gap*. Before the advent and development of the internet, market intelligence companies traditionally relied on classic survey methods to measure customers' perceptions. The measurement of customer-related features started in 1923, with Nielsen at the forefront of measuring competitive sales, which contributed to operationalizing the concept of "market share" and making it a practical marketing management tool. In time, companies increasingly became digital driven, first conducting market surveys online and later generating consumer research and competitive market measurement in the form of analytics testing for actual online behaviors. Competitors such as comScore, Ipsos, GfK, and Kantar have moved in the same direction. For instance, comScore is increasingly becoming a reference source for market and competitive intelligence, thereby allowing companies to benchmark with their competitors, segment their markets, capture relevant market trends, make sense of market data almost in real time, and leverage such data for modeling and forecast purposes. The digital data analytics potential increasingly allows companies to explain what and how competitors are doing in the (digital) marketplace;
5) *The discovery and experimentation knowledge gap*. Despite the reluctance of a number of companies to mine big data in order to discover new patterns and experiments, a few digital data analytics companies have already adopted and are using this technology to support new product development (NPD). For instance, the UK company SoundOut has developed an innovative digital business model to aid its business clients in the entertainment sector to test songs and music tracks before they are released. By administering new music tracks to a panel of 3 million customers in both the US and UK, SoundOut can collect reviews of those tracks and generate predictive analytics data that might be used to either greenlight the production and distribution of a song or

avoid it, thus generating cost savings for its clients. This is certainly the best and highest use that can be made of digital BDA.

While companies differ in their knowledge gaps, BDA stemming from OCRs can help them address such deficits. The most competitive among them are even trying to close knowledge gaps in discovery and experimentation (Davenport, 2014) through significant investments in autonomous analytics leveraging ML (Davenport, 2017).

While studies over the last decade have formulated and tested hypotheses related to the effect of BDA and/or BDAC on organizational performance (Dubey et al, 2019; Ferraris et al, 2019; Fosso Wamba et al, 2020; Gupta & George, 2016), none of them has examined if CGCs possess sufficient BDA and BDAC to innovate their products and services. Indeed, it seems that only a small share of CGCs have developed innovation analytics (Kakatkar et al, 2020) and satisfactory BDAC (Sena et al, 2019). Industry research seems rather to suggest that an increasing number of CGCs rely on the services of professional BDA companies, including providers like Birst, Board International, Domo, Dundas, First Insight, IBM, Logi Analytics, Looker, Microsoft, Microstrategy, Oracle, Pyramid Analytics, Qlick, Salesforce, SAP, SAS, Sisense, SoundOut, Tableau, ThoughtSpot, TIBCO, Yellowfin, etc. (Gartner, 2020). To address this research gap, we formulate and seek to answer the following research question: "As CGCs increasingly rely on professional BDA companies to capture business insights that are critical to innovate their products and services, *how* have they been helped and what types of BDA and BDAC contributed to help them?" To address this research question, we use and explore the revelatory case of the UK-based company SoundOut—a leading provider of BDA for CGCs—and illustrate how it is helping CGCs to conduct BDA-driven innovation.

3. Methods

3.1 Qualitative research design

To bridge the identified knowledge gap, we deployed a qualitative research design based on a case study research technique (Eisenhardt, 1989; Eisenhardt & Graebner, 2007; Yin, 2009). Three reasons justify our choice of this qualitative research method. First, the use of BDA from OCRs for product and business model innovation appears to be a complex and emergent phenomenon, and the related literature is mostly anecdotal and confined to contributions with a practitioner's focus and audiences (Boyd & Crawford, 2012; Bughin et al, 2011; Davenport et al, 2012; Fisher et al., 2012; LaValle et al, 2011; Ohata & Kumar, 2012). Moreover, the notion of BDA for innovation tends to be fluid across time and space (Davenport, 2017). Therefore, a qualitative research design allows obtaining a more nuanced understanding of how BDA and BDAC might help CGCs to undertake innovation and corporate entrepreneurship activities as socially constructed phenomena. Second, while conducting qualitative research, we considered the idea and notion of qualitative pluralism (Cornelissen, 2017), which suggests that adopting qualitative research methods may help to advance theoretical and empirical development, especially when little evidence is available. Combining the case study with a storytelling method is suitable for capturing the nuances and complexity of the questions "how" and "why": how BDA can help CGCs forecast demand for new products in rapidly changing markets in the digital age; and "why" and "to whom" some CGCs outsource BDA for product testing and for innovation purposes. Third, the micro-foundations movement in management studies encourages the adoption of qualitative research methods to generate new insights by considering how individual-level factors aggregate into emergent socio-economic phenomena (Barney & Felin, 2013). Indeed, "appealing to emergence often obfuscates explanation by hiding the actual mechanisms, processes, and actors that lead to the emergent outcome" (Barney & Felin, 2013, p. 147). Accordingly, by focusing on how individual managers active in both BDA companies and CGCs interact in digital experimentation, we also shed light on the emergence of a new way of designing and conducting BDA-driven innovation projects in the digital age.

The overall research approach taken is illustrated in the following methodological considerations. First, as far as the *empirical setting* is concerned, we focused on the consumer goods industry as it is highly reliant on marketing BDA, but scholarly in-depth empirical studies examining innovation BDA are virtually absent. Second, in relation to the *selection of the case*, we decided to focus on a revelatory case (Yin, 2009) as the phenomenon under analysis is emergent and new and since the insights from the real-world case can help explain how CGCs leverage BDA for innovation purposes. More specifically, the case organization is one of the most agile players, having provided innovation BDA to leading CGCs across several countries and continents over almost 15 years. Third, in terms of *data sources*, we collected interview data with both the managers of the case organization and three business clients that used the BDA company for their innovation projects. Accordingly, while the unit of analysis is the organization/s, the level of analysis is the innovation project. In addition to interview data, we also collected contextual information entailing annual reports and other internal and confidential documents such as strategic plans, presentations, and memos. We had preferential access to these documents as one of the researchers has been collaborating with the focal company on a knowledge transfer partnership project for more than two years. Fourth, as far as *data analysis* is concerned, data from documents were triangulated with interview data, and we made validity and reliability considerations, being aware that interview statements can be subject to the personal biases of interviewees related to their responsibilities and roles, as warned by experts in qualitative methods (Creswell, 2014; Yin, 2009). In sections 3.2 and 3.3, we describe the data collection and analysis process.

3.2 Data collection

We adopted convenience sampling to collect our primary data (Pratt, 2009). Data were collected through in-depth semi-structured interviews with a number of interviewees working for the case organization (the UK-based BDA company SoundOut) and for three business clients that are using the BDA company for their innovation projects across three different industries: fashion, retailing, and entertainment and media. The interviews were conducted in two phases. In the first phase, interviews were conducted with the founder/CEO and two members of the executive team (namely the SVP Product and the Data Science Lead) of the UK-based consumer BDA company SoundOut between March 2018 and October 2019. Furthermore, we interviewed three executives of three CGCs that are clients of SoundOut: more specifically, these were the CIOs of the fashion design and online retailing clients, and the SVP Digital Initiatives of the entertainment client. The interviews conducted in the first phase served not only to explore the phenomenon, but also to gain important insights that informed our analyses and subsequent conceptual elaboration. While the interviews conducted in the first phase allowed us to get close to saturation, we decided to conduct additional interviews in 2020. In this second phase, we conducted a second round of interviews with the interviewees involved in the first phase, as well as further interviews with the Vice-President Retail, Project Management Lead, and Development Lead of the BDA company; the Innovation Manager of the fashion design client; the Innovation Manager and Project Management Lead of the online retailing client; and the A&R Manager and the Project Manager of the entertainment client. The interviews conducted in the second phase provided more insights from a larger set of key informants and validated and corroborated the findings from the interview data collected during the first phase. More specifically, based on the knowledge of the interviewees, additional questions were asked to the managers to gain a better understanding of the innovation projects in which SoundOut had been involved. Access to data was possible through our own personal and professional networks. We followed a snowballing approach to contact the interviewees, selecting first those managers and executives that were responsible for BDA in the BDA company and those responsible for the innovation projects carried out by the CGCs. These key trusted referees and informants identified during the first phase were able to indicate further potential interviewees. Table 1 provides an overview of the selected interviewees (and their roles in the respective companies) for this study. Overall, 20 interviews were conducted with 14 key informants: 6 from the BDA company, and 8 from three business clients.

[Insert Table 1 about here]

During the in-depth interviews with the BDA company, we focused on their business models and discussed technical aspects related to their BD technology and how they use it to generate BDA for supporting innovation decisions by their clients—CGCs. Interview questions covered aspects such as how the business model is developed, how business partnerships/relationships with CGCs are created, and what techniques are used to analyze large volumes of OCRs, etc. When interviewing the CGCs, we asked them, among other questions, why and how they were outsourcing data analytics to test new products when they were already closer to end consumers and had significant control over large quantities of data. The interviewees from those companies also answered questions about the characteristics of the innovation projects and the underlying objectives and experiments carried out, and also questions about how their corporate business models are integrated with the business model from the BDA provider. The interviews lasted 55 to 115 minutes and notes were taken by a researcher and a research assistant. During the primary-data collection, the informants interviewed were specialists that were deemed more knowledgeable about the topics being researched. We encouraged our informants to report real-life and real-business stories pertaining to the use of BDA and how it modified their business models (Garud & Giuliani, 2013). Interview transcripts were used to analyze the interview data.

To complement, integrate, and corroborate interview data, we also used archival sources, including annual reports and other internal and confidential documents, such as strategic plans, presentations and memos of the BDA company and internal documents about the innovation projects. These documents allowed us to capture detailed, written information and concrete examples on the partnership activities underlying innovation projects involving SoundOut and its clients. They were employed to gain a richer understanding of the phenomenon investigated.

3.3 Data analysis

Following in the footsteps of qualitative researchers that have developed and used template analysis (Crabtree & Miller, 1999; King, 1998, 2004), we deployed template analysis as it has been found to be particularly suitable for investigating qualitative phenomena in the broad management field, and particularly in information management (Waring & Wainwright, 2008).

In line with template analysis guidelines (King, 1998, 2004), one of the researchers and a research assistant independently read and carried out an initial manual coding of 15% of the printed interview transcripts, either attaching an a priori theme or modifying an existing theme, or even devising a new one. This process allowed them to generate an initial template that included several themes. Subsequently, the themes identified were clustered into a smaller number of higher-order codes in order to describe broader themes in the data. At a later stage, one of the researchers and a research assistant independently read and carried out the manual coding of the remaining 85% of the printed interview transcripts. This resulted in the final template, after revision of the existing themes. During coding, the researcher and research assistant continuously corroborated the insights from analyzing the interviews by constantly triangulating them with the results from analyzing documents. Moreover, during the development of both the first template and the final template, a second researcher (not involved in the coding) conducted a quality check to ensure that assumptions and preconceptions of the other researcher and the research assistant had not systematically biased the analysis. Only then was the final template deployed to generate the findings.

4. Findings: Illustration of the case

4.1 Historical overview of the business, business model pivots, and innovation

The origin. One of the most innovative companies dealing with OCR data analytics in the entertainment and fashion markets, SoundOut was started up in 2007 in Reading (UK) by Mr. David Courtier Dutton, a former city lawyer and corporate finance banker, and veteran of the dot-com boom. The SoundOut platform was developed in parallel with Slicethepie, the first crowdfunding website for unsigned artists. Slicethepie's aim was to be the financing engine for the "new" music industry (see Figure 1). Building on the platform's functions and partnering with social media platforms, it managed to fund about 35 bands, one of which went on to sign for Atlantic Records. However, the underlying business model was not solid enough: the only tiny revenue stream consisted of a small percentage of investments pledged to the band. As clarified by the founder:

> *"This was not enough to make the business profitable... We had not identified our customer segments, though a few users, especially music bands, were benefiting from the crowdfunding mechanisms."*

Certainly, the platform was successful in building a community of loyal music fans. Based on this community, the market insight business SoundOut started offering independent record labels and artists the opportunity to submit songs to Slicethepie's community to generate reviews and market research.

[Insert Figure 1 about here]

Since its foundation, SoundOut has embarked on several business model iterations and pivots that are described sequentially below.

First pivot. In 2009 the company undertook its first pivot by identifying a set of multiple potential customers, including record labels, music publishers, and radio stations. At that time, those companies were struggling to get evidence from the consumers that new music tracks were appealing in the market. This was a major issue as producers of new and emerging artists needed to inform their own decisions on greenlighting products. The SVP Digital Initiatives of a major entertainment and media company said that, at that time, he was challenged by a number of questions:

> *"Should we keep on investing in artists of our portfolio? How should we finance, distribute and sell our products? How should we communicate with potential consumers? These questions needed evidence-based answers and we did not have analytical tools or capabilities to address these questions."*

The observation by Hollywood legend William Goldman that many, if not most, decisions relied on "gut feeling" and that "Nobody-Knows-Anything" remained prominent in all the segments of entertainment and media (Hennig-Thurau & Houston, 2019). This "Nobody-Knows-Anything" mantra was challenged for the first time when SoundOut iterated its business model by offering record labels, music publishers, and radio stations the opportunity to submit songs on its platform and have them tested, based on music fans' reviews drawn from a panel of almost 1 million consumers located in the UK and the US. With the analysis of millions of online reviews, SoundOut started bridging the *online customer satisfaction knowledge gap* as well as the *online customer journey and behavior knowledge* *gap*. Interestingly, the activities carried out with Slicethepie's platform led to an increased penetration rate among music fans. At that time, the platform could collect more than 2.7 million paid reviews and generate analytics data and reports that were subsequently sold to music clients (most of whom were independent entities). SoundOut could therefore have a better understanding of its customers' segments. With revenues stemming mainly from song testing fees from its business customers, the company strengthened its financial status and profitability (see Figure 2).

[Insert Figure 2 about here]

Second pivot. In 2011 SoundOut's CEO, after a meeting with Rob Sisco, the former Head of Entertainment at research intelligence company Nielsen, realized that the company had to focus on two customer segments, namely record labels and radio. Accordingly, the company partnered with all the major recording labels (i.e., Universal Music Group, Sony, Warner Music) to test mainstream tracks for major artists with the aim of identifying the commercial potential of new and unreleased tracks. This second pivot allowed the company not only to consolidate its revenue streams by establishing itself as the *de facto* leader in the generation of predictive market analytics for the music industry, but also to forecast the *digital market potential* of a music track. For instance, in May 2011, SoundOut made an analysis of the album *Born this Way* by Lady Gaga (and subsequently tested products for One Direction and Katy Perry). By focusing on product testing and experimentation, the company was therefore contributing to research and NPD for its clients, enabling them to discover and experiment with new streams and rhythms, to reduce costs, and to optimize inventories based on predictive and prescriptive analytics. Furthermore, clients were in a position to better grasp their competitive intelligence. The company recorded a significant increase in sales and revenues (around £60k/month), also adding other clients in the entertainment and media sector such as CBS, Disney, and the BBC, but it was still lagging behind in terms of profitability.

Third pivot. In 2013 SoundOut undertook a major investment in hardware and software and started providing its application programming interfaces (APIs) to a number of platforms, including the digital music distributor TuneCore and the artist services site ReverbNation. This enabled millions of independent artists to test their music's appeal to the market and the commercial potential of their music in terms of consumers. Moreover, the business scaled up on the vertical segment of the music industry to reach out to an additional customer segment (independent artists) who, despite their inability to pay high commission fees, remained particularly interested in testing their tracks. The availability of APIs allowed SoundOut to further consolidate its business model by adding revenue streams in a scalable way, as the (business) users could use a white label mechanism to integrate the platform functionalities with their own websites.

Fourth pivot. Once SoundOut's CEO validated the platform on the vertical side of the music industry, he started exploring its potential for applications in other sectors such as fashion retailing—where the platform was adapted in 2015 to test different products. Originally it targeted ranging (i.e., the most attractive features of items for consumers and therefore the selected stocks), thus allowing corporate customers to predict future fashion trends, and companies to buy deep on future bestsellers while ensuring eventually lower volumes of unsold stocks. Interestingly, this progressive enlargement of the industry scope brought additional revenues from clients willing to cut inventory costs. The business model of the company therefore became more robust.

Fifth pivot. Based on the high scalability of its platform, SoundOut eventually embraced general retailing in 2015, partnering with companies such as Amazon, Dell, Dixons Carphone, New Look, Warehouse, etc. It also partnered with WGSN, the global market leader in trend forecasting for the fashion industry. Together with WGSN, it launched StyleTrial, which enabled fashion marketers to gain feedback from SoundOut's reviewer community on everything (from one item to a whole collection), which resulted in predictive analytics and reports. By partnering with leading platforms in the fashion industry and other major retail companies, SoundOut had an opportunity to further validate the robustness of its business model. At this stage the company was slightly profitable, but the volume of business was not sufficient to offset the additional cost base and help support the fashion vertical.

Sixth pivot. In 2016 the company embarked on artificial intelligence (AI) technologies and invested significantly in ML to serve its customers. It strengthened its knowledge and competences in both AI and ML by both directly recruiting data scientists and collaborating with field-relevant top-notch PhD students and academics (from Queen Mary's and Goldsmiths). ML enabled SoundOut to map crowdsourced reviews and predictions to actual commercial performance while also accounting for the impact of price, discounting, and other variables. The shift towards AI and ML allowed the company to run trials with retailers such as Argos and Timberland and to start predicting sales on an item by item basis, rather than simply ranking products.

Seventh pivot. In 2017 the company adapted its platform to enable it to ideally deliver software as a service (SaaS) after realizing that a direct sales focus was too resource-intensive (see Figure 3). The strategic focus moved to a *partner model*, which brought over 10 partners from diverse industries to SoundOut and hastened its growth with increased revenues. These multiple partners came from sectors such as entertainment, media, fashion, and retailing—concerning the latter sector, SoundOut recently secured contracts with retailers such as Tesco and e-commerce platforms such as Zalando and Amazon.

[Insert Figure 3 about here]

Eighth pivot. More recently, the company has introduced the concept of consumer enterprise optimization (CEO), thereby enabling businesses to have frictionless consumer input/predictive insights at any point in the supply chain. The ability of SoundOut to provide consumer insight as a “utility” to businesses, together with powerful ML capabilities (to map these insights with business outcomes), offers a tantalizing glimpse into the future of the company’s business and the scale of its ambitions.

Following 12 fundraisings (worth £7m) and 5 (financial) near-death experiences, the company works with 72 shareholders and is taking advantage of a very lean organizational structure (employing 14 people). Its capital is now estimated at £10 million.

The above historical account of the company seems to suggest the following:

- Developing a community of users in the early stages is of paramount importance for a company generating digital data analytics for product testing and demand forecast. The origin of SoundOut as a music crowdfunding platform clearly helped the company to create and develop a community of credible and engaged testers.
- It is rather interesting to note that SoundOut’s original business model mainly served the entertainment vertical, which is the field in which online consumption first emerged and currently represents the leading market for e-commerce in developed economies (comScore, 2018). It was therefore important to work on that vertical segment as a starting point to develop a panel of reliable testers.
- While other companies evolving in the digital space can get online reviews for free (see, for instance, TripAdvisor), the various reviews may not necessarily be accurate and relevant for business decision-making when it comes to quantifying the market potential and setting prices for new products and services (e.g., independent online review websites are awash with fake OCRs).
- Continuous investment in skilled human resources is important to develop the available technologies and acquire competences in new technologies: the move from predictive analytics to AI is certainly emblematic of what other companies working in a similar space are doing.

- A data analytics leader in a specific sector or industry might find it useful to partner with leaders in other industries through cooperative interactions in order to increase its industry scope (see the case of the partnership of SoundOut with WGSN to provide analytics to the fashion industry).
- The way the business model has pivoted over time mirrors the evolution of big data science techniques and related digital technologies.

4.2 Current business model

While describing the unique value proposition of their company, the CEO of SoundOut said:

> *"When you have no data, SoundOut predicts the future. It fuses advanced crowdsourcing, and data science, to predict how new products will perform when they become available to consumers."*

As indicated in section 4.1, the current business model of SoundOut stems from a number of business model iterations and pivots. Overall, it aims to generate and analyze OCRs for completely new products and services, the historic sales data for which are not yet available, as such products/services have not been sold before as they have never been released on the market.

The core value proposition of the company is to provide different forms of predictive digital data analytics to a wide range of businesses and individual customers that have no data about the future and are looking to bridge a few knowledge gaps about (i) online customer satisfaction and customer journey and behavior, and (ii) product testing and NPD. The first knowledge gap is expected to be filled by helping clients to make sense of the *digital market potential* for their offerings, through predictive analytics and demand forecasts. The second gap should be addressed by focusing on discovery and experimentation in order to help clients to reduce costs and optimize inventories based on predictive and prescriptive analytics. As the Data Science Lead explained:

> *"In the space of big data analytics, this company works in a different way. Compared to the [PreviousEmploying_BDACompany] I was working for in London, here we design and build online experiments that are tailored to the specific needs of the client. We can test almost every feature of a new product and service and the quality of the ORs we crowdsource is evaluated using machine learning algorithms that consider almost 50 factors, including relevance, quality of language, plagiarism, etc. Indeed, one of the strength[s] of these analytics is that the quality of the data is the highest I have seen so far…"*

The company has become the global leader in predictive and prescriptive analytics within the music industry, and is consolidating its presence in the fashion industry and in FMCGs. Customer groups are currently segmented in a hybrid way, both by industry and by product, for example retail, advertising (traditional and programmatic), brands (matching brand strategic positioning with creative execution), music, etc.

The company builds on a platform and is positioned as a SaaS white label that contributes to the generation of both OCRs and digital data analytics from OCRs in almost real time.

More than 100,000 OCRs are generated weekly in a *completely controlled environment*; a sort of online experimentation lab used to test new online products and services and generate predictive and prescriptive analytics. To avoid biases in its testing activities that feed into demand forecasting, the platform has developed several mechanisms to ensure the highest quality for both reviewers and reviews. More specifically, it:

- Recruits and signs up online consumers that are paid a fee to write reviews about new products and services, before their release;

- Checks the quality of every review submitted via almost 50 proprietary checks covering relevance, quality of language, plagiarism, etc., and uses ML running across all of these to automatically reject or accept a review;
- Leverages a panel of more than 2.5 million experienced online reviewers covering the US, the UK, and Germany. Reviewers are paid only if the reviews meet certain requirements (in terms of length, period of time taken by the reviewer to write the review, etc.). This check ensures that data can be of the highest quality and that the predictive analytics generated will be as accurate as possible;
- Rewards each and every reviewer for their contribution, unlike many e-commerce websites (e.g., TripAdvisor, Amazon). The platform hosts more than 35 million customer reviews and has paid more than $4.3 million to reviewers so far. However, rewards are paid for each review based on an algorithmic calculation of the "value" of each reviewer in each product category where he/she intervenes;
- Monitors reviewers in real time and normalizes their ratings based on their historic review activity. This ensures that any bias associated with ratings in a skewed or narrow range are still statistically valid;
- Ensures that reviewers' ratings are weighted differently in each product category, based on their historic accuracy in predicting the consolidated group view of their demographic features;
- Analyzes OCRs and calculates an appeal for each product/service tested, based on a methodology of aggregation, weighting, and assessment that is protected by a US patent;
- Employs skilled and talented data scientists, coders and developers, and marketing representatives. Skilled staff is crucial to develop and update the platform and the

associated knowledge and expertise in data analytics, ML, and, more recently, AI. All of them represent crucial distinctive, rare, and hardly imitable assets;

- Performs key activities revolving around data collection from OCRs, storage, processing, cleaning, analysis, and reporting, as well as data science activities and AI. Further activities include data protection and compliance to regulations such as the recently introduced GDPR;
- Engenders customer behavior knowledge as OCR analysis is used not only to accumulate knowledge in relation to customer profiles and aggregate online review patterns, but also to generate micro-segmentation;
- Enhances the market and competitive intelligence of its business clients by carrying out market research about trends through predictive BDA;
- Discovers new trends and carries out new experiments: testing (multimedia) products/services to assess the (digital) market potential; and forecasting demand through a combination of enterprise datasets and ML technologies. In this regard, entrepreneurs and managers are expected to make a crucial decision before launching any new product;
- Shares information with business customers through reports or specific APIs;
- Carries out product innovation as the same platform crowdsources customer evaluations that, in turn, affect the research and development (R&D) processes of its major customers in the entertainment and fashion verticals. Interestingly, it can add value to the R&D activities of third-party companies;
- Enables companies, through its predictive and prescriptive analytics, to make better operational and strategic decisions, and to restructure and rationalize their value chains or several activities of their value chains;
- Embraces AI solutions that might lead to autonomous analytics;

- Introduced the concept of CEO-enabling businesses to access frictionless consumer input/predictive insight at any point in the supply chain. To provide customers with normalized quantitative consumer insights at a speed, scale, and cost to work within a typical customer's workflow, the company introduced the concept of consumer insight as a utility, hardwired into the workflow and combined with ML/analytics, trained on customers' data for near real-time predictions that can optimize the performance of a business;
- Adopts a *partner model* whereby the service is conjointly delivered with other analytics companies (see the project StyleTrial launched in conjunction with WGSN).

Despite the costs related to rewarding reviewers, and maintaining and updating the platforms and staff, the revenues are increasingly relevant and stem from: (1) fees for commission reports delivered to business clients; and (2) subscription fees for using the SaaS.

To summarize, SoundOut's business model is unique and so distinctive, considering the way it generates OCRs and processes and analyzes big data from OCRs. Certainly, the panel of testers represents a point of strength, as exemplified by the SVP Product, who asserted that:

> *"Our real strength is the panel of testers that we have. They are very loyal and do not review only for the tiny amount of money they make out of their reviews, but especially they review because they have a passion for the products they review and they want to engage to help us. The original panel formed back when David convinced me to co-found SoundOut and consists of people that have always reviewed for us, keeping a high standard in their reviewing activity."*

Overall, SoundOut uses a complex and intertwined set of mechanisms to attain its results:

1) Mechanisms to select and monitor reviewers and reviews.

2) Mechanisms and algorithms to reward and incentivize accurate online reviews.
3) Algorithms to analyze OCRs and calculate an appeal score for each product/service tested, based on a methodology of aggregation, weighting, and assessment that is protected by a US patent.
4) Combining SoundOut outputs with client data and using ML to generate client-/item-specific demand predictions.

The aforementioned mechanisms and algorithms ensure that only certified, high-quality, and accurate review data are taken into account to generate predictive and prescriptive data analytics, and that only reliable reviewers are encouraged to leave a review.

4.3 Business relationships with the clients: CGCs

The interview with the selected CGCs—clients of SoundOut—enabled us to raise a number of issues. For instance, the Italian clothing manufacturer, a relatively small company yet significantly internationalized, indicated that the first pilot collaboration with SoundOut in 2015 led to an increased number of interactions. Currently, SoundOut is providing them with predictive analytics on a number of clothing items, and they have launched five new fashion items over the last four years. This means that the manufacturer succeeded in reaching more customers at a faster pace. Moreover, as the US and UK markets are very demanding, they find it reasonable to collaborate with a BDA company that can leverage OCRs from those markets. Furthermore, they still do not possess internal BDAC (for instance, they generate lower-quality data and are not expert in developing predictive algorithms). The CIO of the company recalled a specific project on which they partnered with SoundOut:

> *"We wanted to understand which feature we should prioritize in our new autumn collection of women['s] shoes, and to work out an optimum recommended retail price.*

We exchanged with them information about the reference demographic group to target with our new shoes, and the indicative styles and shapes that we would launch. I had a couple of online meetings with their data science lead and he proposed to run parallel experiments with the same demographic group across two countries, displaying to the customers three features of the shoes' style with an ad hoc algorithm that could change the color. This was useful because we discovered that the potential customers' willingness to pay for the three different styles across the two markets varied by combinations of colors and therefore the color was the key attribute to test conjointly with the new shape."

As for the US entertainment company, it is relatively large and has traditionally invested in digital initiatives since the Napster case. However, they feel that their customers would not respond as neutrally as they do when they administer a new product through SoundOut. Moreover, since their products (songs) are mostly intermediated by other (online) channels such as YouTube or Spotify, they are not so close to their consumers. While they collect and store a relatively large amount of data, these are not the "right" data to foresee the behavior of a new song or an emerging artist in the market. In addition, they still lack a larger panel of potential testers. The A&R Manager commented on a project that was mentioned by the SVP Digital Initiatives of the company during the first phase of the interviews:

"Well, for that project the band [a globally renowned band] was about to release a new album. However, there was a kind of 'standoff' between the band and us, as to which track to release as a first single. As we could not agree, we decided to test the whole album with SoundOut to gather more data to help resolve the matter. SoundOut conducted an experiment to test all 12 of the album tracks monadically in the UK and US markets with hundreds of consumers for each track. After normalizing and weighting the responses, each of the 12 tracks was rated on a 0–100% percentile scale

both by country and by demographic, thus enabling a comparison. We called in SoundOut to meet with the band and presented the results of the research. They managed to persuade the band that their chosen single was most unlikely to be commercially successful. As a result, we were able to release the chosen single that went on to achieve a No. 1 position in the UK charts."

The German e-commerce company can rely on a large volume of OCRs that are very frequently left by the customers on the e-commerce platform. However, those reviews are mostly written by reviewers that are already actual customers and relate only to existing products/services transacted on the website. Most of the reviews are written after a verified purchase. This implies that OCRs are functional to the core business of the platforms, which consists of economic transactions. However, they are not used to test new products or forecast the demand of products that must be launched on the market. As such, the e-commerce company needs a BDA partner to carry out those experiments that will be necessary to understand if a new design or product could sell more. Furthermore, the company prefers to use the analytics of actual customers to understand more about their actual behaviors and thus generate transactional value, rather than engage with potential customers to test a new product.

5. Discussion

5.1 Key findings and contributions

We make several contributions to the literature on BDA, notably contributing to the nascent research stream revolving around innovation analytics (Kakatkar et al, 2020), by arguing that innovation BDA and innovation BDAC can make a difference for CGC innovation initiatives.

First, we find that CGCs can partner with BDA companies to generate knowledge and insights into the market potential of a new product (not yet launched on the market) by setting up online experiments in a digital experimentation lab—under the guise of a controlled online platform—where product testing and market research are carried out for product and service innovation, demand forecast, and market potential prediction. This finding illustrates key trends in practice and research: in practice, an increasing number of companies is engaging with innovation BDA that are often co-created with BDA companies specialized in digital experiments; in research, scholars should quickly move beyond a generic characterization and operationalization of BDA and BDAC and build on the nascent stream of innovation analytics (Kakatkar et al, 2020) not only to identify the antecedents of effective *innovation* BDA and *innovation* BDAC, but also to gain a better understanding of the effect of *innovation* BDA and *innovation* BDAC on organizational performance, thus adding to the extant literature exploring the relationship between BDA and performance (Nam et al, 2019; Williams, 2016), and BDAC and performance (Ferraris et al, 2019; Mikalef et al, 2019a, 2020; Rialti et al, 2019; Srinivasan & Swink, 2018; Fosso Wamba et al, 2020). More precisely, it seems that a specific emphasis should be put on innovation performance—as a specific dimension of organizational performance—and the holistic process by which innovation BDA can support innovation initiatives in digital settings.

Second, conducting effective digital experiments leading to meaningful innovation BDA requires a few *preconditions* and a well-defined *holistic process*. As far as pre-conditions are concerned, there should be: (1) good communication flows between the innovation managers of the client company and the BDA provider to ensure that the business questions to address and business hypotheses to test are accurately identified, shaped, and formulated; (2) seamless conduction of the experiment in a protected digital environment (accessible only to testers and not the wider online audience); (3) crowdsourcing of online reviews in relatively high volumes by leveraging a large and representative consumer panel; (4) constant quality control of online reviews and reviewers by means of ad hoc algorithms (possibly patented);

and (5) a proficient use of data mining techniques (including advanced ML) to discover relevant patterns in the data in relation to satisfaction with and/or preference for a new-to-the-market product: this implies that *innovation* BDAC—an evolution of generic BDAC (Gupta and George, 2016)—should be in place. As far as the *holistic process* is concerned, there is a well-defined series of steps that should be followed. In the first step, a digital experiment is set up in a controlled online environment (the underlying assumption of the experiment being that the tested consumer product will be well received by online consumers). In the second step, the consumer product is then administered to an online panel of consumers/testers. In the third step, OCRs are crowdsourced in an isolated and controlled experimental environment. These OCRs feed a database of previous reviews (from the same or other consumers) that allows a double-check of the quality of both the review (comparatively with other reviews written by the same reviewer in the past for other products or reviews written by others for the same product) and the reviewers. In the fourth step, OCR data flows in the form of digital data streams (Pigni et al., 2016). In the fifth step, by deploying innovation BDAC, the OCR data are first pre-processed and processed, and later analyzed. The final step entails the generation of reports for innovation BDA (these can be in the form of reports and graphs also using visualization tools). This holistic process, which we define as the CGCI BDA framework, is illustrated in Figure 4:

[Insert Figure 4 about here]

Third, and interestingly, the CGCI BDA framework entails the presence of both resources in line with the resource-based view of the firm (Barney, 1991, 2001) and organizational capabilities, be they dynamic or not (Amit & Schoemaker, 1993; Eisenhardt & Martin, 2000; McEvily & Zaheer, 1999; Teece & Pisano, 1994). Compared to the conceptualization of BDAC developed by Gupta and George (2016), who identified tangible, human, and intangible resources, our framework is distinctive for different reasons. Concerning tangible resources, data are the by-products of an experiment and are crowdsourced in a protected experimental environment (thus allowing the protection of innovation BDA and insights). With regard to human resources, whereas the managerial skills contribute to identifying the right business question and hypotheses to test (and are therefore related to the business acumen of the innovation manager), the technical skills make a difference in implementing the online experiment and getting the results through processing and analysis of the data. As far as intangible resources are concerned, a data-driven culture is by construction present in the BDA company, even though the customer might display a less pronounced data-driven culture. However, lasting collaboration on a series of innovation projects might improve the data-driven culture of the CGC.

Fourth, the CGCI BDA framework shows that the way data are generated and managed does not necessarily follow a circular data life cycle (Blazquez & Domenech, 2018); it is mainly a directional process with a clear direction pointing towards a specific product innovation initiative. This is illustrated in Figure 5:

[Insert Figure 5 about here]

Fifth, the cases analyzed provide evidence supporting the idea that CGCs increasingly tend to either outsource or co-create the holistic process underlying the generation of innovation BDA for new product testing and innovation initiatives of digital BDA companies. Typically, outsourcing happens due to a number of reasons: (1) digital BDA companies are perceived as organizations having superior innovation BDAC; (2) the CGC has not carried out any previous project leveraging innovation BDA; (3) online consumers perceive digital BDA companies as independent from companies dealing in consumer goods. In other words, consumers are not given information about the CGC or brand whose product is meant for testing, and this puts them in a more comfortable position; (4) CGCs, as also shown in e-commerce websites, believe that OCRs from a third-party environment (and not from their own website) are less biased and more objective and reliable when it comes to product testing; (5) while outsourcing BDA for innovation, CGCs also outsource the liabilities of recruiting reliable respondents as well as those of protecting the data; (6) the data generated by a BDA company are often perceived as being of higher quality compared to a BDA from a CGC: the core business of a BDA is nothing but BDA, coupled with the fact that BDA companies typically have a multi-year track record in analyzing big data from consumers, and have therefore developed robust proprietary algorithms; and (7) even medium and large CGCs mainly rely on analytics for their current and previous customers, but they do not necessarily have direct access to larger panels of potential consumers: if they want to know more about potential customers, they are expected to go further with their existing customers' base by exploring different niche markets/market segments.

That said, and increasingly, it seems that the more CGCs experience innovation BDA and develop an internal data-driven culture, the more their approach towards the process of innovation BDA generation will be in line with co-creation. Accordingly, the development of a partnership between the BDA company and the CGC would be conducive to higher extracted value from innovation BDA.

5.2 Theoretical implications

Several theoretical implications emerge from this work. First, based on the analysis of the revelatory case study, we contribute to the literature with a novel framework, the CGCI framework, that illuminates how BDA companies are technically helping CGCs to innovate. This framework blends innovation BDA and innovation BDAC. If we employ the definition of BDAC developed by Gupta and George (2016), the CGCI framework shows that: (1) the tangible resources are highly protected as innovation activity is very sensitive and both the BDA company and its clients need to avoid knowledge spillover; (2) managerial skills make a difference in identifying the right business question and the right hypotheses to test (and they sit originally in the CGC and later are also developed in the BDA company, once the partnership evolves over time), while technical skills make a difference in implementing the online experiment and getting the results through processing and analysis of the data (technical skills sit mainly in the BDA company but can be partially acquired by the CGC); and (3) while a data-driven culture is by construction present in the BDA company, the customer might display a less pronounced data-driven culture and might, therefore, use BDA to make different decisions.

Second, data quality within the CGCI framework is of paramount importance, and several CGCs are increasingly relying on BDA from companies whose core business is BDA: this seems to enrich the perspective that a quality dominant logic in BDA (Fosso Wamba et al., 2018) should be in place to enhance not only firms' performance, but also firms' innovation BDAC. Indeed, "poor data quality or ineffective data governance is a key challenge for big data" (Fosso Wamba et al, 2015, p. 244) and for their transformation into BDA that are so critical for innovation.

Third, the BDA companies that are more relevant for innovation purposes are those that are able to generate innovation analytics by online crowdsourcing reviews in a digital experimental lab. This contributes to shift the focus of extant BDA empirical studies that have mainly dealt with predictive finance and customer analytics (Batistič & van der Laken, 2019) and extend it to innovation analytics (Kakatkar et al, 2020), which is an advancement over mere R&D analytics (Dremel et al., 2020).

Fourth, while digital experiments represent a powerful mechanism to innovate for CGCs, it seems that data security and data governance mechanisms should be increasingly shaped at the inter-organizational level and along the value system and chain, or broadly, along the technology delivery system (TDS) (Huang et al, 2018). This implies that while the supplier of BDA is responsible for the storage and security of data and BDA, it falls upon the client of BDA to protect BDA in the wider BD industry (Kwon et al., 2015).

Fifth, we illustrate that CGCs often lack advanced BDA and BDAC and therefore rely on external BDA companies for innovation projects. Interestingly, this does not involve complete outsourcing of innovation activities and projects, as cross-organizational innovation teams are built to encompass both managers in the CGC and managers in the BDA company. This implies that BDAC are co-created between the client and the provider of BDA.

Sixth, and related to the previous point, an ecosystem of BDA enterprises specializing in innovation BDA is emerging, as they can create transformational value that is sold to CGCs. This phenomenon needs to be considered when mapping digital business ecosystems (Kwon et al, 2015; Tsujimoto et al, 2018), notably by digital entrepreneurs in their value propositions (Nambisan, 2017). This means that the distributed locus of agency underpinning digital entrepreneurship should become a priority in both digital management and entrepreneurship research streams. This is in line with the findings by Caputo et al., 2019, who concluded that: "in the world of Big Data, innovation, technology transfer, collaborative approaches, and the contribution of human resources have a direct impact on a company's economic performance" (p. 6).

5.3 Practical implications

This study offers several practical implications targeting three different audiences/stakeholders: (1) innovation managers and CTOs of CGCs, (2) BDA companies' managers, and (3) CIOs and data scientists.

As far as innovation managers are concerned, mounting competition between CGCs should lead them to more frequently and proactively look for predictive *innovation analytics*. This implies that they should either build innovation BDAC in-house or liaise with BDA providers. Large companies, firms already equipped with powerful digital technologies, firms endowed with digital skills, and organizations endowed with a data culture, might find it easier to build innovation BDAC in-house. However, those companies that lack a BDA culture or possess few technological resources should reconsider their innovation strategies by hiring BDA companies and/or outsourcing BDA for some of their innovation projects and initiatives.

Second, the decision about whether or not to outsource innovation BDA, and how, should be preceded by a cost/benefit analysis by the innovation managers. If CGCs outsource BDA, they can benefit from: leveraging the huge capital, technological, and human investments already made by the BDA company; accessing the state-of-the-art BDAC and specialized skills; and adopting a technological solution that has already been trialed. These benefits, of course, come at a cost, represented typically by the fees charged by the BDA company. Contrary to what has been observed in recent literature (Dubey et al, 2019), the return on BDA investment is almost immediate and materializes as soon as the innovation analytics reach the innovation managers' dashboard, and a decision is made based on those analytics. This is particularly true for consumer goods with short life cycles like fashion, entertainment, and music products. On the other hand, while developing BDA internally could allow more control over the BDA, protection of data, and potentially more flexibility, it would require a relevant investment in technology and specialized skills, often yielding suboptimal results compared to what professional BDA companies can achieve.

Third, and related to the previous point, CGCs should move a step beyond a mere make-or-buy decision and consider if the specific innovation project or the market context requires a purchase of innovation analytics or rather to shape a partnership with the providers of BDA. In markets where products are subject to fads and quick modifications of consumers' preferences, building partnerships with BDA companies should be pursued. More specifically, innovation managers should make sure that BDA companies integrate SaaS seamlessly in their innovation processes and systems to improve their capability to innovate at a faster pace, and should form an inter-organizational innovation team. The latter is critical for shaping the appropriate hypotheses and assumptions for experimentation and testing, as the CGC internal innovation team possesses the business acumen necessary to shape and formulate the appropriate business question(s) and hypotheses (and understand the assumptions), whereas the BDA team is endowed with the right digital skills to understand the feasibility of (and implement) the digital experiment. Accordingly, a conjoint innovation structure, bringing together the CGC's innovation team and the BDA company team allocated to the innovation project, is a necessary condition to co-create innovation value effectively; it plays a critical role in accurately setting up the experiment and interpreting the experiment's results and the related BDA and insights related to the specific market where the experiment took place. For this to happen, the CGC needs to trust the BDA provider, communicate with it frequently to gain an understanding of the provider's capabilities and limitations, be open to a data-driven culture (Gupta and George, 2016), and learn to apply the knowledge generated by the BDA provider.

Fourth, in shaping a partnership with the providers of BDA, innovation managers should be ready to build long-term relationships allowing them to support innovation projects over time and involve multiple functions. Indeed, innovation culture and data-driven culture should go hand in hand across multiple functions (Gupta and George, 2016), thus resulting in the breakdown of knowledge silos and allowing all the CGC functions to embrace an innovation and BDA-driven culture. In practice, this could be implemented by bringing together the internal innovation team managers across different functions, for all of them to potentially benefit from the insights into potential customers and new markets, and to translate them into a business model.

Fifth, CTOs of CGCs should always be part of innovation projects for which a BDA provider is selected. This is critical to guarantee high levels of security and control when the

BDA providers integrate SaaS seamlessly (through APIs) into the information systems of the CGC.

As for the BDA companies' managers, they will have to seize several opportunities and face challenges associated with their operations. First, they should articulate their unique value proposition. For instance, before the establishment of SoundOut, most of the managerial decisions in the music industry were based on "gut feeling". This state of affairs has changed, and music recording companies know that the mitigation of innovation risks related to the release of a new track/album requires regular testing through SoundOut. This is becoming increasingly pervasive for both short life cycle products (like fashion and entertainment products) and products with a relatively longer life cycle (like consumer electronics). Overall, BDA companies are well equipped to reduce innovation risks across an increasing number of industries and, therefore, we expect them to be recruited by CGCs to work on a growing number of innovation projects.

Second, instead of limiting themselves to recognizing the important role of BDA by industry players (Dresner Advisory Services, 2018), BDA companies should keep on building an appropriate BDA mindset, skillset, and toolset, as well as an internal culture of experimentation (Thomke, 2003, 2020). Building such a mindset, skillset, toolset, and experimentation culture might empower BDA companies to educate their prospective clients on the potential of BDA to inform innovation decisions, including business model innovation through iteration and pivoting. This way, BDA companies will be able to attract more CGCs to their client portfolio in the near future.

Third, BDA companies should increase their investment in technologies such as IoT and virtual and augmented reality to increase the realism of their experiments and to capture more holistically customers' needs. At the same time, they should increasingly invest in security not only to comply with regulatory frameworks, but also to improve the confidence of their clients. This is key to avoiding pernicious data leaks that could compromise the reputation of data providers and affect online consumers' privacy, security, and, ultimately, their trust and behaviors.

Fourth, and last, BDA companies should strike a balance between their interests and those of their clients. Unlike transactional platforms, BDA companies need adequate value creation mechanisms not only for their customers, but also for themselves. For instance, since the crowdsourcing of UGC is not free of charge, SoundOut has to handle high costs, which means that other monetization mechanisms are required to guarantee profitability in the medium term. More generally, BDA companies should devise new ways to appropriate value by means of the strongest monetization mechanisms (for example, a share of the value they generate for their clients) to ensure the long-term viability of their business model.

Concerning the CIOs and data scientists of BDA companies and CGCs, their tendency to operate in a silo and to interact least with other business functions and organizational units is an issue to tackle. A more appropriate and effective use of BDA requires data scientists to communicate effectively with all the functional areas and, in particular, with innovation/R&D and marketing services (for example, with CTOs, CMOs, and COOs). In addition, as the boundaries across traditional industries become increasingly blurred, data scientists working in BDA companies should gain a holistic approach of business processes and adopt analytical solutions for placing a new product in contiguous markets. Finally, as Cambridge Analytica's case clearly suggests, BDA companies are expected to avoid data leaks, as their effects might be catastrophic. CIOs and data scientists are therefore required to establish effective and efficient data security systems.

6. Conclusions

CGCs in several industries and sectors are looking for new ways to speed up the pace of their innovation initiatives, without losing their ability to be quick to respond to highly competitive, uncertain, and dynamic global markets. Meanwhile, a new ecosystem of BDA companies seems to have the potential to support product and business model innovation by means of BDA. As illustrated by the SoundOut case study, BDA companies are able to create and deliver strategic and transformational value to client CGCs by crowdsourcing ad hoc OCRs via well-designed digital experiments. When combined with ML techniques, data analytics (both predictive and prescriptive) is relevant to bridging several knowledge gaps, including forecasting demand for new products and determining the online market potential of new goods.

Interestingly, CGCs should keep partnering with BDA companies to ensure that BDA becomes a natural extension of the company's innovation infrastructure (which will add to its internal assets and competences). By outsourcing BDA to third-party companies or by interacting with them, CGCs can trigger transformative changes that can trump entrepreneurial activities.

With regard to the limitations of this work, the first one is related to the adopted approach. While the case study approach is suitable for exploring the phenomenon under scrutiny in depth (Eisenhardt & Graebner, 2007; Yin, 2009), the potential drawback resides in that the generalization of the findings is possible only if they are being validated using a larger sample of cases in different geographical settings. However, since BDA companies operate in digital environments and have multiple clients based in different countries, the geographical limitation is not a relevant issue. Second, even though several of the findings and conclusions of this work might apply to other sectors as well, we focused primarily on the role that BDA can play for CGCs, with one exemplary and revelatory case study: SoundOut. Therefore, similar longitudinal case studies on innovative companies catering their innovation to additional firms and industries are welcome to enrich the conceptualization of the proposed CGCI BDA framework.

Future research might follow three different directions. First, it might enrich the conceptualization of the CGCI framework in light of the literature on embedded analytics (O'Donovan et al., 2018), as fog computing might become an increasingly interesting tool to protect innovation BDA that are extremely precious for the client company. Second, scholars might build on the framework we have proposed, as well as on the resource-based theory-based conceptualization of BDAC (Gupta and George, 2016), to strengthen the conceptualization of an innovation BDAC, which might significantly expand the nascent field of innovation analytics (Kakatkar et al, 2020) and R&D analytics (Dremel et al, 2020). Third, and in the light of the increasing reliance of firms on BDA, it might make sense to compare competing models based on the transaction cost economics (Williamson, 1979) and service-dominant logic (Vargo & Lusch, 2008), so as to understand to what extent CGCs should rely on external providers of BDA, build partnerships to co-create BDA for their innovation initiatives, or rather develop an internal infrastructure for innovation BDA. Last, it might be interesting to explore how the emerging ecosystem of BDA companies could be better described by leveraging the literature on digital entrepreneurship (Nambisan, 2017) and entrepreneurial ecosystems.

References

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics, 182*, 113–131.

Amit, R., & Schoemaker, P. J. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, *14*(1), 33–46.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17,* 99–120.

Barney, J., & Felin, T. (2013). What are microfoundations? *The Academy of Management Perspectives*, *27*(2), 138–155.

Basuroy, S., Chatterjee, S., & Ravid, A. (2003). How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing, 67,* 103–117.

Batistič, S., & van der Laken, P. (2019). History, evolution and future of big data and analytics: A bibliometric analysis of its relationship to performance in organizations. *British Journal of Management, 30*(2), 229–251.

Bean, R. (2018). How big data and AI are driving business innovation in 2018. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/how-big-data-and-ai-are-driving-business-innovation-in-2018/ Accessed 7 April 2019.

Beath, C., Becerra-Fernandez, I., Ross, J., & Short, J. (2012). Finding value in the information explosion. *MIT Sloan Management Review, 53*(4), 18–20.

Blazquez, D., & Domenech, J. (2018). Big data sources and methods for social and economic analyses. *Technological Forecasting & Social Change, 130,* 99–113.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society, 15*(5), 662–679.

Brandenburger, A.M. & Nalebuff, B.J. (1996). *Co-optetition*. New York: Doubleday.

Bughin, J., Livingston, J., & Marwaha, S. (2011). Seizing the potential of 'big data'. *McKinsey Quarterly, 4*, 103–109.

Caputo, F., Mazzoleni, A., Pellicelli, A. C., & Muller, J. (2019). Over the mask of innovation management in the world of big data. *Journal of Business Research*, in press.

Chen, H., Chiang, R., & Storey, V. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4) (December 2012), 1165–1188.

Chen, Y., Liu, Y., & Zhang, J. (2012). When do third-party product reviews affect firm value and what can firms do? The case of media critics and professional movie reviews. *Journal of Marketing, 76*, 116–134.

Chiang L.-L.L., & Yang C.-S. (2018). Does country-of-origin brand personality generate retail customer lifetime value? A big data analytics approach. *Technological Forecasting & Social Change*, *130*, 177–187.

Chintagunta, P., Gopinath, S., & Venkataraman, S. (2010). The effect of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science, 29*, 944–957.

Chong, A., Li, B., Ngai, E., & Ch'ng, E. (2016). Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. *International Journal of Operations and Production Management, 36*(4), 358–383.

Clemons, E. K., Gao, G. G., & Hitt, L. M. (2006). When online reviews meet a study of the hyperdifferentiation: Craft beer industry. *Journal of Management Information Systems, 23*(2), 149–171.

Cornelissen, J. P. (2017). Preserving theoretical divergence in management research: Why the explanatory potential of qualitative research should be harnessed rather than suppressed. *Journal of Management Studies, 54*(3), 368–383.

Cox, M., & Ellsworth, D. (1997). Managing big data for scientific visualization, ACM Siggraph, *MRJ/NASA Ames Research Center, 5*, 1–17.

Crabtree B. F., & Miller, W. L. (1999). *Doing qualitative research*. Sage.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed method approaches*. (4th ed.). Thousand Oaks, CA: Sage Publications.

Cui, G., Lui, H.-K., & Guo, X. (2012). The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce, 17*(1), 39–58.

D'Aveni, R. A. (1994). *Hypercompetition: Managing the dynamics of strategic maneuvering*. New York: Free Press.

D'Aveni, R. A. (1995). *Hypercompetitive rivalries*. New York: Free Press.

D'Aveni, R., Dagnino, G. B., & Smith, K. (2010). The age of temporary advantage. *Strategic Management Journal, 31*(13), 1371–1385.

Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review, 84*(1) (January), 98–107.

Davenport, T. H. (2014). How strategists use "big data" to support internal business decisions, discovery and production. *Strategy & Leadership, 42*(4), 45–50.

Davenport, T. H. (2017). How analytics have changed in the last 10 years. *Harvard Business Review*, June.

Davenport, T. H., Barth, P., & Bean, R. (2012). How “big data is different”. *MIT Sloan Management Review*, Fall. https://sloanreview.mit.edu/article/how-big-data-is-different/ Accessed 12.10.2019

Dean, J. (2014). *Big data, data mining and machine learning: Value creation for business leaders and practitioners*. New York: Wiley.

Dremel, C., Herterich, M. M., Wulf, J., & Vom Brocke, J. (2020). Actualizing big data analytics affordances: A revelatory case study. *Information & Management, 57*(1), 103–121.

Dresner Advisory Services (2018). *2018 big data analytics market study*. Wisdom of Crowds® series of research.

Duan, W., Gu, B., & Whinston, A. (2008). Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems, 45*(4), 1007–1016.

Dubey, R., Gunasekaran, A., Childe, S. J., Blome, C., & Papadopoulos, T. (2019). Big data and predictive analytics and manufacturing performance: Integrating institutional theory, resource-based view and big data culture. *British Journal of Management, 30*(2), 341–361.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532–550.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *The Academy of Management Journal*, *50*(1), 25–32.

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal, 21*(10–11), 1105–1121.

Ferraris, A., Mazzoleni, A., Devalle, A., & Couturier, J. (2019). Big data analytics capabilities and knowledge management: Impact on firm performance. *Management Decision, 57*(8), 1923–1936.

Fisher, D., DeLine, R., Czerwinski, M., & Drucker, S. (2012). Interactions with big data analytics. *Interactions*, *19*(3), 50.

Forman, C., & Ghose, A. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information System Research, 19*(3), 291–313.

Fosso Wamba, S., Akter, S., & de Bourmont, M. (2018). Quality dominant logic in big data analytics and firm performance. *Business Process Management Journal*, *25*(3), 512–532.

Fosso Wamba, S., 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.

Fosso Wamba, S., Dubey, R., Gunasekaran, A., & Akter, S. (2020). The performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism. *International Journal of Production Economics,* 107498.

Fosso Wamba, S., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356–365.

Gartner. Magic quadrant for analytics and business intelligence platforms. (2020). https://www.gartner.com/en/documents/3980852/magic-quadrant-for-analytics-and-business-intelligence-p Accessed 8 April 2020.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, *59*(5), 1493–1507.

Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. *Journal of Business Research, 108,* 147–162.

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

Hennig-Thurau, T., Gwinner, K. P., Walsh, W., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, *18*(1), 38–52.

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Systems, 57*, 42–53.

Huang, Y., Porter, A. L., Cunningham, S. W., Robinson, D. K. R., Liu, J., & Zhu, D. (2018). A technology delivery system for characterizing the supply side of technology emergence: Illustrated for big data & analytics. *Technological Forecasting & Social Change, 130*, 165–176.

Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information System Research*, *15*(2), 194–210.

Jun, S.-P., Park, D.-H., & Yeom, J. (2014). The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference. *Technological Forecasting & Social Change, 86*, 237–253.

Kakatkar, C., Bilgram, V., & Füller, J. (2020). Innovation analytics: Leveraging artificial intelligence in the innovation process. *Business Horizons, 63*(2), 171–181.

Kapoor, R., & Lee, J. M. (2013). Coordinating and competing in ecosystems: How organizational forms shape new technology investments. *Strategic Management Journal, 34*, 274–296.

Kim, K., Park, O.-J., Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting & Social Change, 123*, 362–369.

King, N. (1998). Template analysis. In G. Symon, & C. Cassell (Eds.), *Qualitative methods and analysis in organisational research: A practical guide*. London: Sage.

King, N. (2004). Using templates in the thematic analysis of text. In C. Cassell, & G. Symon (Eds.), *Essential guide to qualitative methods in organizational research*. London: Sage.

Kwon, T., Kwak, J. H., & Kim, K. (2015). A study on the establishment of policies for the activation of a big data industry and prioritization of policies: Lessons from Korea. *Technological Forecasting & Social Change, 96*, 144–152.

Laney, D. 3D data management: Controlling data volume, velocity and variety. META Group Research Note, 6. (2001). https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf Accessed 4 April 2019.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review, 52*(2), 21–32.

Liu, X., Shin, H., & Burns, A. C. (2019). Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing. *Journal of Business Research*, in press.

Mariani, M. M., Baggio, R., Fuchs, M., & Höpken, W. (2018). Business intelligence and big data in hospitality and tourism: A systematic literature review. *International Journal of Contemporary Hospitality Management, 30*(12), 3514–3554.

Mariani, M. M., Predvoditeleva, M. (2019). How do online reviewers' cultural traits and perceived experience influence hotel online ratings? An empirical analysis of the Muscovite hotel sector. *International Journal of Contemporary Hospitality Management*, 31(12), 4543-4573.

Mariani, M.M., Borghi, M., & Kazakov, S. (2019). The role of language in the online evaluation of hospitality service encounters: An empirical study. *International Journal of Hospitality Management*, 78, 50-58.

Mariani, M.M., Di Fatta, G., & Di Felice, M. (2019). Understanding Customer Satisfaction with Services by leveraging Big Data: the Role of Services Attributes and Consumers' Cultural Background, *IEEE Access*, 7,8580523, 8195-8208.

Mariani, M. M., Borghi, M., & Gretzel, U. (2019). Online reviews: differences by submission device. *Tourism Management*, 70, 295-298.

Mariani, M. (2019). Big Data and analytics in tourism and hospitality: a perspective article, *Tourism Review*, 75(1), 299-303.

Mariani, M., & Borghi, M. (2019). Industry 4.0: A bibliometric review of its managerial intellectual structure and potential evolution in the service industries. *Technological Forecasting & Social Change, 149,* 119752, https://doi.org/10.1016/j.techfore.2019.119752.

Mariani, M.M., & Visani, F. (2019). Embedding eWOM into efficiency DEA modelling: an application to the hospitality sector, *International Journal of Hospitality Management*, 80, 1-12.

Mariani, M., & Borghi, M. (2020). Online review helpfulness and firms' financial performance: An empirical study in a service industry. *International Journal of Electronic Commerce*, in press.

Mariani, M., Borghi, M.,& Okumus, F. (2020). Unravelling the effects of cultural differences in the online appraisal of hospitality and tourism services, *International Journal of Hospitality Management*, 90, 102606. https://doi.org/10.1016/j.ijhm.2020.102606

McEvily, B., & Zaheer, A. (1999). Bridging ties: A source of firm heterogeneity in competitive capabilities. *Strategic Management Journal*, *20*(12), 1133–1156.

Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L., & Canhoto, A. (2018). Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research*, *93*, 67–78.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019a). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research, 98*, 261–276.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019b). Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. *British Journal of Management, 30*(2), 272–298.

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management, 57*(2), 103169.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34*(1), 185–200.

Nam, D., Lee, J., & Lee, H. (2019). Business analytics use in CRM: A nomological net from IT competence to CRM performance. *International Journal of Information Management, 45*, 233–245.

Nambisan, S. (2017). Digital entrepreneurship: Toward a digital technology perspective of entrepreneurship. *Entrepreneurship Theory and Practice, 41*(6), 1029-1055.

O'Donovan, P., Gallagher, C., Bruton, K., & O'Sullivan, D. T. J. (2018). A fog computing industrial cyber-physical system for embedded low-latency machine learning Industry 4.0 applications. *Manufacturing Letters, 15*, 139–142.

Ohata, M., & Kumar, A. (2012). Big data: A boon to business intelligence. *Financial Executive, 28*(7), 63–64.

Pigni, F., Piccoli, G., & Watson, R. (2016). Digital data streams: Creating value from the real-time flow of big data. *California Management Review, 58*(3), 5–25.

Pratt, M. G. (2009). From the editors: For the lack of a boilerplate: Tips on writing up (and reviewing) qualitative research. *Academy of Management Journal*, *52*(5), 856–862.

Rialti, R., Zollo, L., Ferraris, A., & Alon, I. (2019). Big data analytics capabilities and performance: Evidence from a moderated multi-mediation model. *Technological Forecasting & Social Change, 149*, 119781.

Sena, V., Bhaumik, S., Sengupta, A., & Demirbag, M. (2019). Big data and performance: What can management research tell us? *British Journal of Management, 30*(2), 219–228.

Shen, B., Choi, T.-M., & Chan, H.-L. (2017). Selling green first or not? A Bayesian analysis with service levels and environmental impact considerations in the big data era. *Technological Forecasting & Social Change,* 144: 412-420.

Srinivasan, R., & Swink, M. (2018). An investigation of visibility and flexibility as complements to supply chain analytics: An organizational information processing theory perspective. *Production and Operations Management*, *27*(10), 1849–1867.

Teece, D., & Pisano, G. (1994). The dynamic capabilities of firms: An introduction. *Industrial and Corporate Change, 3*(3), 537–556.

Thomke, S. H. (2003). *Experimentation matters: Unlocking the potential of new technologies for innovation*. Boston: Harvard Business Review Press.

Thomke, S. H. (2020). Building a culture of experimentation. *Harvard Business Review*, March–April.

Tsujimoto, M., Kajikawa, Y., Tomita, J., & Matsumoto, Y. (2018). A review of the ecosystem concept — Towards coherent ecosystem design. *Technological Forecasting & Social Change, 136*, 49–58.

Vargo, S. L., & Lusch, R. F. (2008). Service-dominant logic: Continuing the evolution. *Journal of the Academy of Marketing Science, 36*(1), 1–10.

Wang, Y., Kung, L. A., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting & Social Change*, *126*, 3–13.

Waring, T., & Wainwright, D. (2008). Issues and challenges in the use of template analysis: Two comparative case studies from the field. *The Electronic Journal of Business Research Methods, 6*(1), 85–94.

Williams, S. (2016). Business intelligence in the era of big data and cognitive business. In S. Williams (Ed.), *Business intelligence strategy and big data analytics* (pp. 27–50). Boston: Morgan Kaufmann.

Williamson, O. E. (1979). Transaction-cost economics: The governance of contractual relations. *The Journal of Law and Economics, 22*(2), 233–261.

Yin, R. K. (2009). *Case study research: Design and methods*. Sage Publications.

You, Y., Vadakkepatt, G. G., & Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing, 79*, 19–39.

Table 1. An overview of the companies/people (interviewees) who participated in this study

Case	*Industry*	*Business description*	*Headquartered*	*Interviewee's position/s (number of interviews)*
A	BDA	Digital BDA	UK	Founder/CEO, SVP Product, Vice-President Retail, Project Management Lead, Development Lead, and Data Science Lead (N=9)
B	Fashion design	Manufacturer of clothing/fashion	Italy	CIO, Innovation Manager (N=3)
C	Online retailer	E-commerce company	Germany	CIO, Innovation Manager, Project Management Lead (N=4)
D	Entertainment and media	Music recording	US	SVP Digital Initiatives, A&R Manager, Project Manager (N=4)

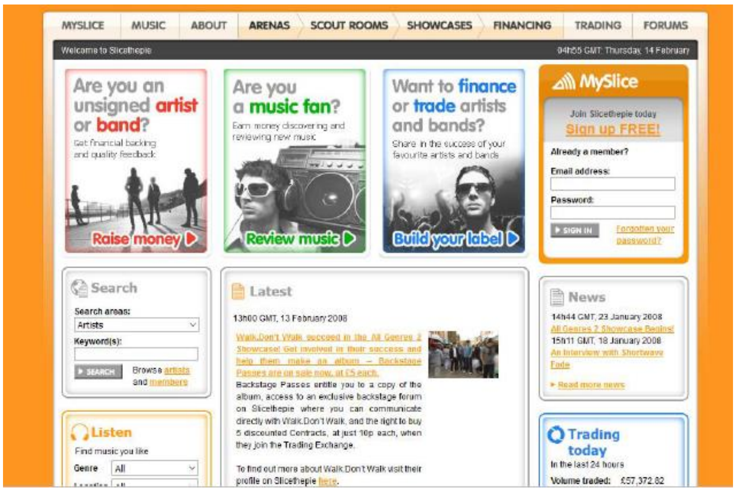

Figure 1. Slicethepie as a precursor of SoundOut, 2007

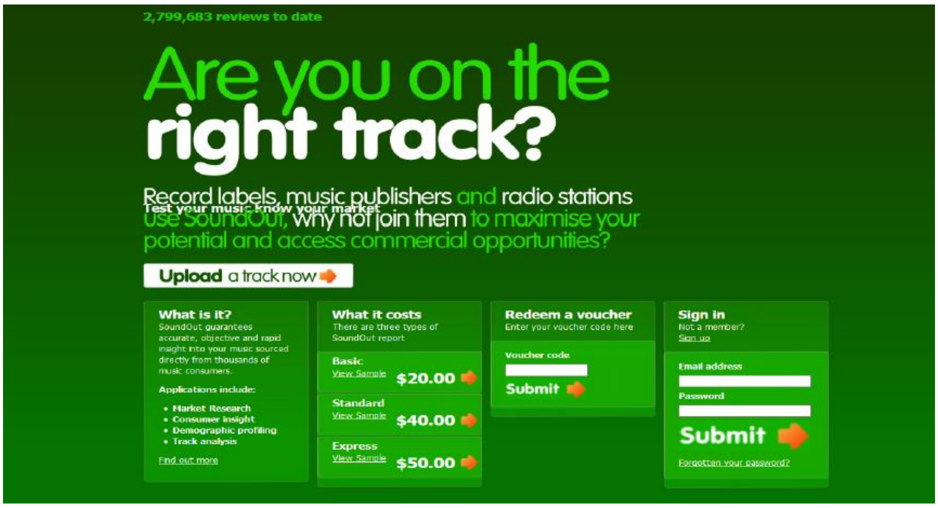

Figure 2. First pivot of SoundOut: Testing sound tracks for independent artists, record labels, music publishers, and radio stations, 2009

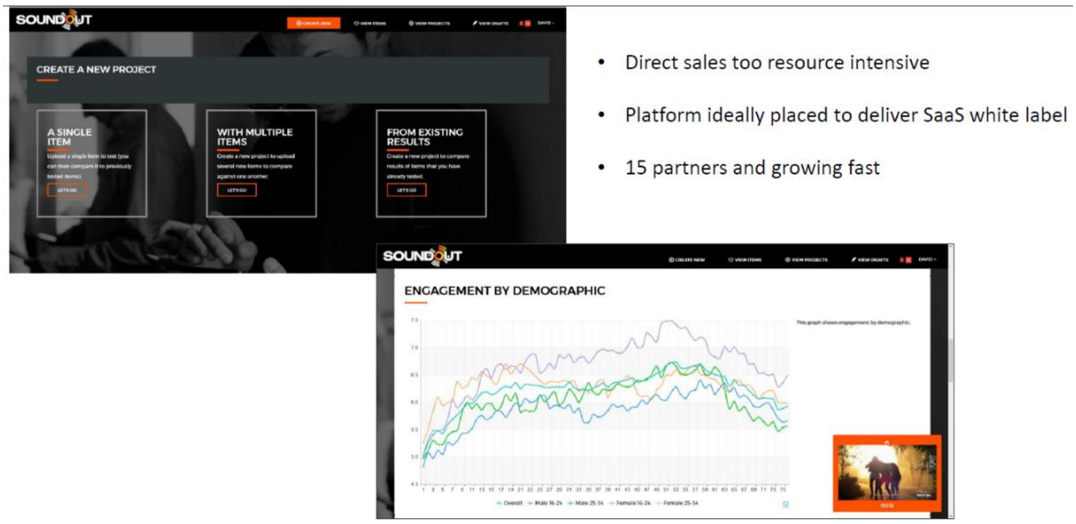

Figure 3. Seventh pivot of SoundOut: Delivering software as a service, 2017

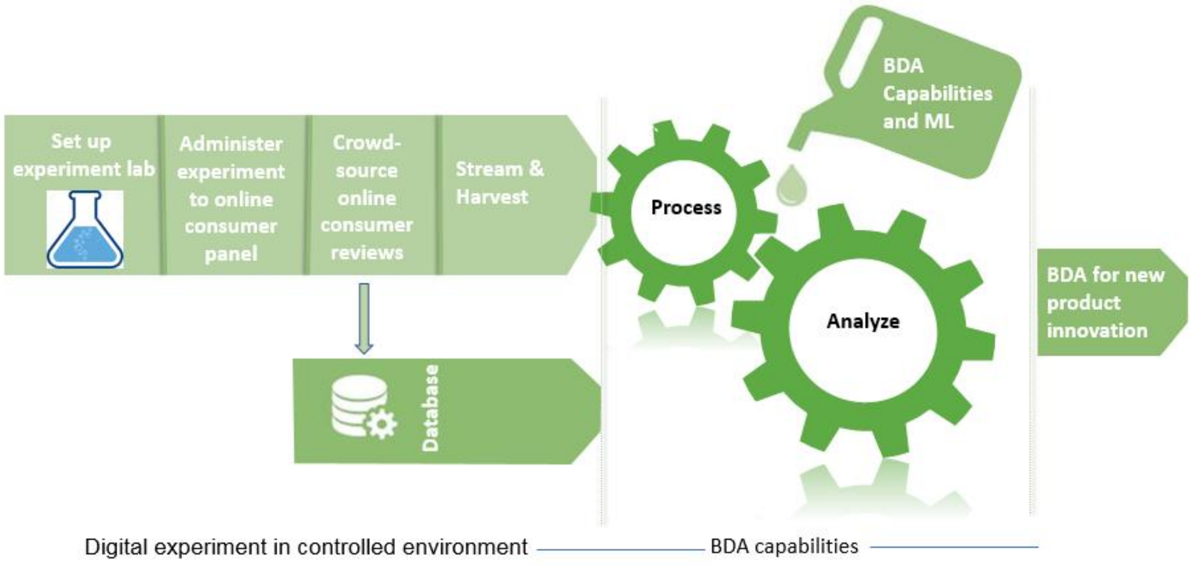

Figure 4. CGCI BDA framework

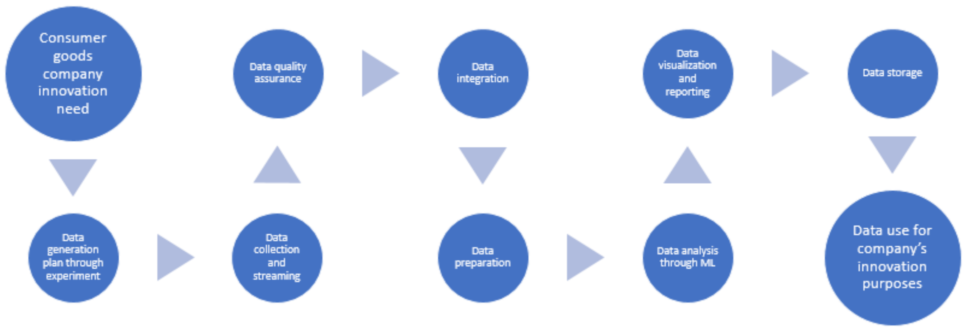

Figure 5. Data flow for CGCI