Paper to be presented at the DRUID Academy Conference 2019
at Aalborg University, Denmark
January 16-18, 2019

Is Big Data Overrated? The underestimated innovation challenges of BD management

Vella Verónica Somoza Sanchez
SDU
CI2M
vella@sam.sdu.dk

Peter Schneider-Kamp
University of Southern Denmark
IMADA
petersk@imada.sdu.dk

René Chester Goduscheit
SDU
CI2M
rene@sam.sdu.dk

Wolfgang Gerstlberger
University of Southern Denmark
Department of Marketing and Management
woge@sam.sdu.dk

Abstract

The exponential growth of data all over the world in the last 15 years (McAfee & Brynjolfsson, 2013), has originated the field of studies of big data (BD). The literature on BD has studied the potential benefits associated to the analysis of large quantities of data and how this can be used to improve the organization; from its operations to the development of new products (for example: Gandomi & Haider, 2015; McAfee & Brynjolfsson, 2013). Furthermore, data has been portrayed as the key to solve some of the current societal problems (Patterson, 2017). Additionally, the current big disruptions of society and business have a digital nature, as for example the creation and use of artificial intelligence in organizations. However, for this data to be translated into value, it must be first transformed into knowledge (Zins, 2007). To this end, it is important to take into account that this data production is directly connected to the humans, as it is mainly produced by them or about them (McAfee & Brynjolfsson, 2013). Therefore, ignoring the human component of the current level of data may only result on missing out on its potential. There is an important effort to be made as to consider the human factor in the analysis of data being currently produced since it has the potential to be used as to develop new products and services that may contribute to solve some of these important social problems. To this end, we present a new experimental method based on BD analysis that integrates the

human perception for new product development (NPD). This way, taking to a higher degree advantage of the data potential. This method is called Human-in-the-loop (HITL), a model with human interaction used for modeling and simulation for the cinematographic and gaming industry (Rothrock & Narayanan, 2011). Here is adapted as an experimental method integrating human perception into the analysis of BD. It is composed by two dynamic loops of data analysis and a method of connection. The first loop regards the human interaction through the introduction of human behavior in the form of experiments or ethnographic data. The second loop regards computer-based analysis of structured data impulse introducing new data sets. The mixed method of digital anthropology serves as the bridge to link structured and unstructured data types.

However, as the case shows, there are risks to the implementation of BD in the organization; as producing better answers for organizational problems brings inevitably new questions that require unforeseen resources (George, Osinga, Lavie, & Scott, 2016). Through the take-aways from this case study, as well as the literature on the topic, we propose a theoretical contribution using the framework of George et al. (2016). We establish that given the current technology of data analysis there are challenges for data-driven innovations depending on the data scope and granularity that each case regards. More specifically, we state that the current BD technology is unable to confront simultaneously high granularity (variety) and scope (volume) for analysis. Therefore, when companies try to give answer to the new questions that BD brings along, they enter the unknown, represented by what we define as the ?gravitational sink' of BD. This is defined as such, as much of the solutions explored by organizations will fail into this sink. However, due to the potential of disruption of the remaining technologies promised by the technology-driven literature (Habtay, 2012), organizations are attracted to its exploration. Additionally, we argue that ambidextrous organizations are in a better position to undergo the challenges of the implementation of BD technologies as exploration and exploitation should happen as a dualism due to the dynamic nature of the BD technologies. Therefore, the research objective of this paper can be summarized in the following question: What is the role of data granularity and scope in the management of digital disruption and innovation for BD?

Finally, we provide a guide for practitioners, showing which standardized technologies available are the best fit for each data scenario. This way allowing for certain level of exploration; yet avoiding the gravitational sink. For example, when the scope of the data regards a small volume with no variety, it can be handled with statistics, or manual data processing. Once the data scope increases, BD processing can help handling large volume of data. Tools for business intelligence can effectively analyze this volume in absence of variety. Additionally, when the variety increases, tools as NLP or image recognition provide a solution. Once these technologies are in play, the resources allocated for data analysis grow due to the problems presented by Del Vecchio et al. (2018): need of skilled analysts, increasing data storage and processing cost, security risks, etc. When the data presents increased variety and a small volume, traditional ethnographic methods handle data thickness in the unstructured data (Wang 2016). Nevertheless, once the data volume increases, digital ethnography is required to analyze larger data sets. These methods, help explore new questions, but none of them fully answer high volume and variety data related problems. This exploration of the unknown represented by the gravitational sink presents high level of challenges due to its dynamism and continue need of exploration. However, it also has the potential for long lasting disruption.

George, G. et al. (2016) 'Big Data and Data Science Methods for Management Research', Academy of Management Journal, 59(5), pp. 1493-1507.

Habtay, S. R. (2012) 'A Firm-Level Analysis on the Relative Difference between Technology-Driven and Market-Driven Disruptive Business Model Innovations', Creativity and Innovation Management, 21(3), pp. 290-303.

Jansen, J. J. P., Van Den Bosch, F. A. J. and Volberda, H. W. (2006) 'Exploratory Innovation, Exploitative Innovation, and Performance: Effects of Organizational Antecedents and Environmental Moderators', Management Science, 52(11), pp. 1661-1674.

Del Vecchio, P. et al. (2018) 'Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges', Creativity and Innovation Management, 27(1), pp. 6-22.

Zins, C. 2007. Conceptual approaches for defining data, information, and knowledge. Journal of the American Society for Information Science and Technology, 58(4): 479-493.

Is Big Data Overrated? The underestimated innovation challenges of BD management

Vella Somoza Sánchez*

Department of Marketing and Management. University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark.

E-mail: vella@sam.sdu.dk

Peter Schneider-Kamp

Department of Mathematics and Computer Science. University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark.

E-mail: petersk@imada.sdu.dk

René Chester Goduscheit

Department of Marketing and Management. University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark.

E-mail rene@sam.sdu.dk

Wolfgang Gerstlberger

Department of Marketing and Management. University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark.

E-mail: woge@sam.sdu.dk

* Corresponding author

The literature has focused on the potential benefits of big data analysis (for example: Gandomi & Haider, 2015). However, the challenges associated to its management have been neglected. This research seeks to contribute to the discussion towards the challenges that the implementation of the big data technologies brings for the organization. This way, we discuss how there is no standardized technologies to deal with data with high scope and granularity. Therefore, companies must develop their own tailor-made solution able to deal with their data. However, when the level of data entails high scope and granularity we define the existence of the gravitational sink, where many of the tailor-made technologies will fail due to the unforeseen challenges associated to these dynamic technologies. However, due to the potential of disruption of the remaining technologies, organizations are attracted to its exploration.

Keywords: Big data management; experimental mixed methods; big data challenges; ambidexterity.

1 Introduction

The exponential growth of data all over the world in the last 15 years (McAfee & Brynjolfsson, 2013), has originated the field of studies of big data (BD). The literature on BD has studied the potential benefits associated to the analysis of large quantities of data and how this can be used to improve the organization; from its operations to the development of new products (for example: Gandomi & Haider, 2015; McAfee & Brynjolfsson, 2013). Furthermore, data has been portrayed as the key to solve some of the current societal problems (Patterson, 2017). Additionally, the current big disruptions of society and business have a digital nature, as for example the creation and use of artificial intelligence in organizations. However, for this data to be translated into value, it must be first transformed into knowledge (Zins, 2007). To this end, it is important to take into account that this data production is directly connected to the humans, as it is mainly produced by them or about them (McAfee & Brynjolfsson, 2013). Therefore, ignoring the human component of the current level of data may only result on missing out on its potential. There is an important effort to be made as to consider the human factor in the analysis of data being currently produced since it has the potential to be used as to develop new products and services that may contribute to solve some of these important social problems. To this end, we present a new experimental method based on BD analysis that integrates the human perception for new product development (NPD). This way, taking to a higher degree advantage of the data potential. However, as the following case shows, there are risks to the implementation of BD in the organization; as producing better answers for organizational problems brings inevitably new questions that require unforeseen resources (George, Osinga, Lavie, & Scott, 2016). Through the take-aways from this case study, as well as the literature on the topic, we propose a theoretical contribution using the framework of George et al. (2016). We establish that given the current technology of data analysis there are challenges for data-driven innovations depending on the data scope and granularity that each case regards. More specifically, we state that the current BD technology is unable to confront simultaneously high granularity (variety) and scope (volume) for analysis. Therefore, when companies try to give answer to the new questions that BD brings along, they enter the unknown, represented by what we define as the ‘gravitational sink’ of BD. This is defined as such, as much of the solutions explored by organizations will fail into this sink. However, due to the potential of disruption of the remaining technologies promised by the technology-driven literature (Habtay, 2012), organizations are attracted to its exploration. Additionally, we argue that ambidextrous organizations are in a better position to undergo the challenges of the implementation of BD technologies as exploration and exploitation should happen as a dualism due to the dynamic nature of the BD technologies. Finally, we provide a guide for practitioners in figure 5, showing which standardized technologies available are the best fit for each data scenario. This way allowing for certain level of exploration; yet avoiding the gravitational sink. For example, when dealing with variety and high volume of data, tools like natural language processing are a good solution. The same way, when the data presents high variety and some volume, digital ethnographic techniques are of help.

Therefore, the research objective of this paper can be summarized in the following question:

What is the role of data granularity and scope in the management of digital disruption and innovation for BD?

2 Theoretical framework

Big data definition

The definition of BD stands for the collection of data that includes the following characteristics; volume, velocity and variety (McAfee & Brynjolfsson, 2013). Volume regards the storage space for the datasets analyzed (Del Vecchio et al., 2018), velocity refers to the rate in which the data grows

(McAfee & Brynjolfsson, 2013), as well as how quickly it is shared and stored (Del Vecchio et al., 2018). Variety regards to the different types of data that can be encountered; this way, structured numeric data coexist with unstructured data (as pictures, videos, text, etc.) not easily tabulated and structured into variables (McAfee & Brynjolfsson, 2013). These characteristics are known as the 3Vs of BD and are often used in the literature to define the concept of BD (McAfee & Brynjolfsson, 2013). The literature has also attended to the many possibilities of BD (Gandomi & Haider, 2015), applications (Park, Schwartz, & Eichstaedt, 2015), associated methods and tools (George *et al.*, 2016), and ethical problems (Salganik, 2017). It has become an important topic due to its disruptiveness for decision making with potential for impacting society and individuals (Jin, Wah, Cheng, & Wang, 2015). The academic research on the topic, has explored its many possibilities from each discipline's point of view; in management studies, BD presented a revolutionary way to analyze larger datasets to provide better answers to current questions and explore new frontiers (George et al., 2016). In other disciplines, the focus attended to other alternatives and implications, as for example for the field of study (Feldman, Kenney and Lissoni, 2015; Mayer-Schonberger & Cukier, 2013), or the possibilities of the combination of the different fields (Salganick, 2017).

Scope and granularity of BD

In management theory, George *et al.*, (2016) state that the previous 3Vs of BD can serve to theory production. However, the characteristic of velocity is usually left to explore to computer science studies (George et al., 2016). Here, this definitory dimension is considered as an intrinsic part of BD, playing an important role in the implications that the implementation of this technology brings for the organization. George *et al.*, (2016) define the scope and granularity of big data in relation to the remaining 2Vs as follows: first, moving from samples of data to nearly data populations greatly improves the scope of the theories potentially produced (volume). Second, using different data types (variety), theories can shift towards holistic answers considering for example, effects in different research groups and therefore, improving the results' granularity (George et al., 2016). Therefore, for these authors, the volume and variety in the data are the key to improve the scope and granularity of management theory. This improvement regards not just producing better answers to already known problems in management studies, but also new questions exploring the unknown. More specifically, George *et al.*, (2016) find that only data with high granularity and scope have the potential to produce these new questions developing management theory. However, even though there is a great potential to the development of new theory though the exploration of the data scope and granularity, these authors do not evaluate the challenges of this exploration, and the tension that it may bring to the current activities of the organization.

The challenges of BD in the literature

Reviewing the literature of the challenges associated to BD, two sets of problems are identified; technical problems associated to the management of data and challenges transforming data into knowledge for improvement of organizational competitiveness (Del Vecchio et al., 2018).

The first set of problems focuses on discerning what data is important for the organization, making sure that the analysis of the data is possible, acquiring the necessary resources, and using the lections learnt through the data to transform the organizational operations (McKinsey & Company, 2013). Additionally, these authors state challenges regarding personal privacy and data security, reliability and veracity. Blazquez and Domenech (2018), based on Kitchin (2014), see as a major challenge the loss of meaning when translating unstructured data into structured variables. These authors establish that social theory and contextual knowledge is needed to explain the formation of patterns in unstructured data. Furthermore, computer analysis is not ready yet to fully automatize these explanations (Blazquez & Domenech, 2018). Gandomi and Haider, (2015) discuss the potential for application of BD tools and techniques and a set of problems for predictive analysis of BD. These are the need for sophisticated statistical techniques, the possibility of noise (error) accumulation in the calculations due to the sheer size of the data, the presence of spurious correlations (false correlations between variables), and incidental endogeneity (independence of the variables/predictors from the error term is presupposed wrongly undermining the validity of the model). George *et al.*, (2016), gather these challenges into five items: Data access and collection, where security and privacy as part of this process, not generalizable to the overall analysis. Data storage, where oversimplification to save storing space may lead to problems of reproducibility (George et al., 2016). Data processing related to the transformation of unstructured data as text, and emotions into numerical variables subject of statistical analysis (George et al., 2016). The loss of data thickness in the sense of Kitchin (2014) is overlooked by these authors. Data analysis using techniques as Ridge, lasso, regression trees or bootstrapping to solve the analytics problems described by Gandomi & Haider, (2015). Data reporting and visualization for evaluation of the decisions made and presentation of results to different audiences.

The second set of challenges focus on how to transform data into knowledge for organizational improvement (Del Vecchio et al., 2018). In this set, three additional Vs related to BD have been defined. These characteristics are Veracity, Variability and Value (Del Vecchio et al., 2018). Veracity relates to the reliability of the data interpretation (Del Vecchio et al., 2018). Variability studies problems in connection to the continuous changes of ever growing datasets and value relates to data-driven decision making and scalability problems of unstructured data (Del Vecchio et al., 2018). Nevertheless, these challenges do not specify how the implementation of BD technologies affects the organization and its activities.

The ambidextrous organization

The ambidexterity literature studies the challenges of exploration and exploitation for the organization. Moreover, the ambidextrous organization is defined as the one exploring new frontiers, adapting to the ever changing environment and overrunning competition, that at the same time ensures the stability necessary to exploit current competences (Jansen, Van Den Bosch, & Volberda, 2006 based on Levinthal and March, 1993). The literature on ambidexterity regards three approaches towards achieving ambidexterity: The first approach focuses on cycles of exploration and exploitation non-concurrent but cyclical (Turner, Swart, & Maylor, 2013). This is known as temporal ambidexterity (Turner et al., 2013). The second way to establish this equilibrium is through the role of a common manager for different business units dedicated to either exploitation or exploration, where, even if both activities happen at the same time, they come from different business structures (O'Reilly & Tushman, 2004). This dualism is known as structural ambidexterity (Turner et al., 2013). The third possibility relies on the employees capacity to move towards exploration or exploitation considering the needs of the organizational context, without relying on a formal strategy (Cao, Gedajlovic, & Zhang, 2009). This is known as the contextual approach (Turner et al., 2013).

3 The case

The case selected is a small company employing less than 20 employees in Odense, Denmark. This company uses BD to improve its business offerings: websites comparing banking products. More specifically, small loans, credit cards, unemployment unions, dating sites, tv services, and broadband services. They function as intermediaries between the users and the final service providers. The company has 71 webpages in seven different languages and operates in 7 countries; US, Sweden, Spain, Poland, Norway, Finland, and Denmark. Where, in terms of returns, Sweden, Denmark, Spain, and Norway are its main markets. The company's revenue comes from two activities; the subscriptions and purchases to the final provider's services. Currently, the company uses a BD architecture tracking its websites users' IP to store information into variables. These variables compose a real time system used to track some of the characteristics of the users. This data is used to select the order in which the products appear on the company's websites, for search engine optimization (SEO) and marketing tools for Google advertising (SEM). This companies uses BD analysis for deciding the order of the offerings appearing in its websites. Using the theory from George *et al.* (2016), these (improved) answers, coming from data analysis lead towards new questions, explaining the decision of the company of pursuing personized website experiences for each visitor.

The company explores and exploits in cyclical phases through separated structures: the IT department works on exploration of new technologies and methods and the marketing department exploits the BD infrastructure through SEO and SEM. Therefore, the marketing department acts as the follower of the developments of the IT department. Even though the organization seeks an equilibrium between the expenses associated to exploration and the revenue generated by the marketing department, there is no common manager orchestrating a strategy for both business units. Therefore, taking this into account we classify the organization under the temporal ambidexterity umbrella.

This case has relevance for to two reasons: first, it allows the application of the experimental method of human-in-the-loop (HITL) that requires their BD infrastructure, and second, it serves to illustrate the challenges of BD implementation for the organization.

4 Methodology

The HITL approach

The concept of the HITL relates to a model in which there is human interaction (Rothrock and Narayanan, 2011). This concept has been used in the computer science field for modeling and simulation for the cinematographic and gaming industry (Rothrock and Narayanan, 2011). In the current case it is adapted as an experimental method inspired by this concept adapted to integrate human perception in the analysis of BD. Figure 1 shows the method and its three elements; two dynamic loops of data analysis and a method of connection.

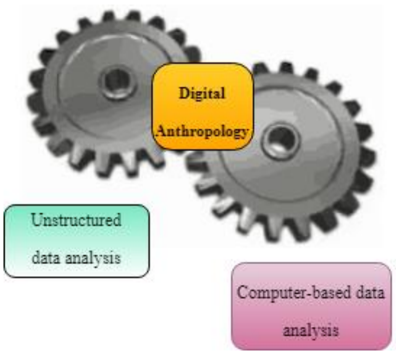


Figure 1 The-Human-in-the-loop approach

The first loop regards the human interaction in the model. This loop moves through the introduction of human behavior in the form of experiments or ethnographic data. The second loop regards computer-based analysis of structured data. This loop impulses through the introduction of new data sets. The method of intersection that links together both loops is the mixed method of digital anthropology (Somoza Sánchez, Giacalone, & Goduscheit, 2018). The mixed method approach serves as the bridge to link both structured and un-structured data types. It works by using the behavioral data as a form of supervised learning for the analysis of structured data. The behavioral information in the model serves to understand the local culture produced in the settings of the particular case (Horst & Miller, 2012) and that knowledge is then used to holistically analyze the dataset studied. Therefore, the both dynamic loops connected by the method of digital anthropology are interdependent as to produce the same goal; the holistic analysis of big data sets.

Quantitative data collection

The company tracks its website users' IP to store information about them. The organization's database is organized in two layers. The first layer, summarized in table 1, is used in the company as a hypothetical fingerprint of the website's users allowing the identification of some of the user's characteristics as their location, their internet browser, the device from which the query was launched, and, in case that these users come from a google campaign; the keyword bringing them into the website visited. The second layer of the database states the relationships between the previous descriptive variables and their seasonal trends. The two layers of data serve to automatically select which products should be first on the list of offerings that the users see on the different websites. Due to data protection restrictions, the variables and algorithm used by the company in the second layer are not further described. These two layers of data where introduced as the second loop in the HITL model, using the insights from the behavioral data, described in the upcoming section, as the first loop.

Table 1 Fingerprint data in time series available (From August 2016 to December 2017)

Variable	*Definition*
website_id	Internal reference to identify the website
website_name	Textual name of the website
country_id	Internal reference to the users' country
country_name	Textual name of the country
source_channel_id	Internal reference to the marketing channel used for advertising
channel_name	Textual name of the marketing channel
visit_date	It states a visit to a website not being necessarily the first visit of this user

click_date	First time that the user clicks on any tag or button on the webpage
product_id	Internal reference to the product in the company's database
product_name	Textual name of product
device_id	Internal reference to the device from which the request was made
device_name	Device from which the request was made
nr_clicks	Number of clicks per product
revenue	Total revenue per product, per day (visit date) in Danish crowns.

Source: Own work

Qualitative data collection

In this HITL model, once the structured data was collected, human insights give purpose to the data; in this case we used the behavior of the website's users. This data was introduced in the HITL model through a design thinking (DT) approach. The reasons behind this choice are that DT is been proposed as a good tool for rapid prototyping in the literature (Beckman & Barry, 2007) and for the advantages that it offers for studying behavioral data in the context of website design (Plattner, Meinel, & Leifer, 2011). More specifically, seeking to preserve the dynamism of the HITL approach, a DT workaround method was introduced because it offers an iterative process between the abstract ideation process and a concrete resulting design through a continuous prototyping process (Beckman & Barry, 2007). The first stage a DT process focuses on the observation of the context in which the user operates to understand how products are being used and how to make sense of this behavior (Beckman & Barry, 2007). The second stage seeks to explain what is important for the user experience (Beckman & Barry, 2007). The third stage tries to understand how the value proposition changes and improves after the inclusion of the new framework created in the previous stage. The last stage regards the concrete features that the new product or service must have to meet this new value proposition (Beckman & Barry, 2007). Throughout this phase, the learning process is completed should be used as an iterative process seeking for disagreements and challenges, as to continuously reframe and improve ideas (Beckman & Barry, 2007).

Digital Anthropology using software tools

Virtual ethnography was used as base to conduct the DT process. This is an observation method to study Internet data (Kozinets, 2013). This method is adequate for two reasons: The product at hand, small personal loans, is a sensitive product therefore, recruiting respondents proved a difficult task. Second, the method is a good fit with the method of connection of this experimental HITL approach; the method of Digital anthropology (DA). The particularities of DA are three: study of the source of data, use of triangulation to understand the network of relationships present in the data, and attention to each case (Somoza Sánchez et al., 2018). DA introduces the local culture produced in the website of observation as an element to consider for the data analysis (Somoza Sánchez et al., 2018). This is especially relevant to produce framework in the DT process, since the way the users interact with the architecture of the website is an important cue to understand how and why the users behave in a specific manner. DA is used as to find relevant stories to create the DT frameworks. For doing so, the authors watched 1017 videos produced by the software tools employed; Hotjar and Smartlook. These software tools record the users' visits in videos, in which the activity of each user can be followed individually. Additionally, the software tools summarize the interaction recorded in the videos in activity maps, used in the analysis. During the observation of those videos, the context was considered as to better understand the navigation of the websites and to follow the principles of DA of using all the sources of data available for data triangulation (Somoza Sánchez et al., 2018).

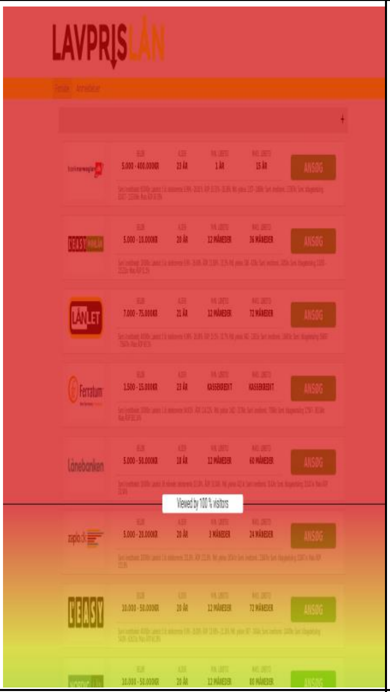
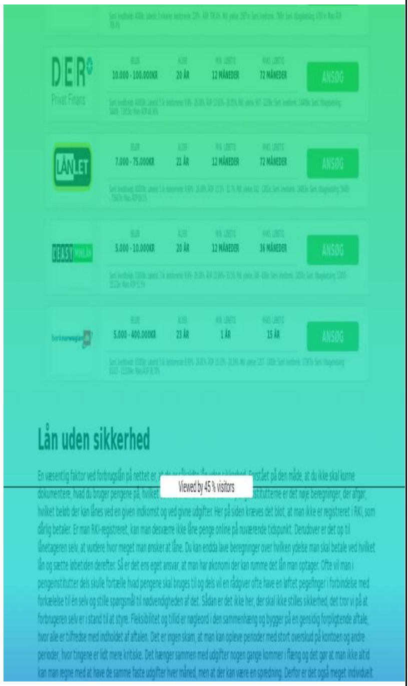
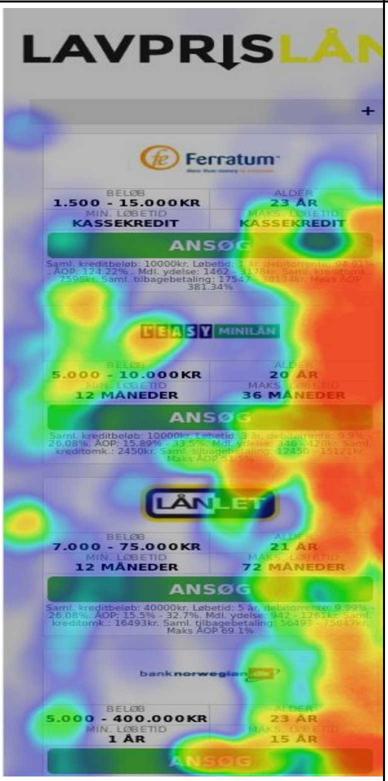
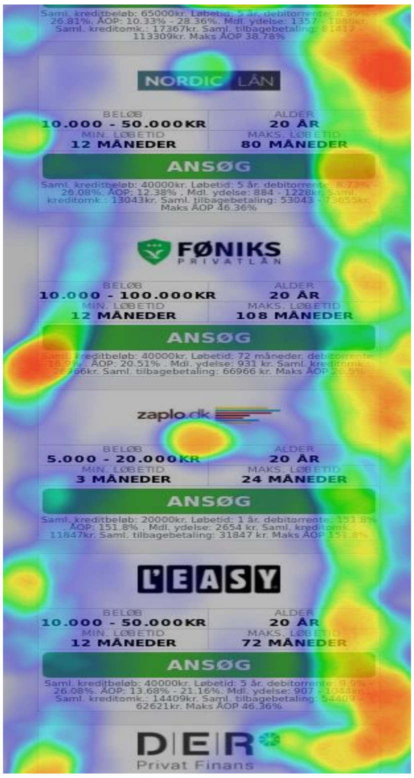
5 Analysis

The objective seek by the organization through the experimental HITL approach was to produce personalized website designs for each website visitor. For doing this, a DT process was carried out producing as result a first set of website designs prototypes. This process will be presented in the first part of the analysis. Afterwards, we use the failure of the case to explain the theoretical contribution in the second section of this analysis, were the gravitational sink is presented in Figure 5.

The DT process

The analysis of the data was separated attending to the user's device for accessing the websites; desktop computer or mobile phone/tablet. The data collected through the software tools was divided into three variables: scrolling behavior, clicking behavior and mouse movement tracking. Heat maps, as table 2 shows, were used to interpret the data, together with the observation of the videos and the feedback from the company's employees. This tables is an example of the visualizations of those heat maps. More specifically, they show two of the variables in which the data was gathered; the scrolling behavior for the visitors using desktop computers and the clicking behavior for mobile phone users. Where the red colors show the parts of the website with higher user activity.

Table 2 Scrolling and clicking behavior for desktop and mobile phone devices

View by 100% of the visitors	View by 60% of the visitors	View by 45% of the visitors
		
Top part of the website for mobile phones	Middle part of the website for mobile phones	Bottom part of the website for mobile phones
		

Source: Own work

The results of the observations showed that the users payed attention to the top of the sites and to the products with larger and most saturated icons. For mobile phone devices this behavior was more accentuated, where only the 25% of the users scrolled to the complete bottom. This implies, that the users did not scroll enough as to read the terms of refund, or any other text on the bottom of the websites. Around 40% of the users entered the sites but did not engage in any other activity. The tapping behavior of the mobile phone devices showed that the users were right handed and used their thump to scroll and click. Additionally, some visitors used the mouse or their thump as a help to read the text of the products primarily from the top of the sites. These observations were used as to understand the users' behavior or their lack of interaction, as well as what elements could have been missing for those users. These stories were collected in the DT framework represented in Figure 2. This figure explains the two profiles in which the users of the websites selected were classified; the students and the window-shoppers. The first profile grouped the users who were experts on these types of products and spent longer time on the sites. Those users with a desktop device usually subscribed to the newsletter and were considered older since they were using the cursor to go over the text of the different products. These users often clicked back and forth between different products, therefore, a comparison system, were several products could be confronted was suggested as the new value proposition. The window-shoppers profile corresponds to those users connecting to the sites through social media channels; specifically, from Facebook campaigns or advertisings. These users did not expend much time on the sites, often leaving without interaction. This behavior was interpreted as a lack of interest on the websites. Therefore, one of the main objectives of the prototypes focused on dragging attention. For doing so, a principle of simplicity was decided including two elements; an explanation of how the process worked and a search function. The final prototypes for each profile can be observed in figure 3 and 4, where the new value proposition for each profile is reflected in the new designs. Figure 3 shows the students' prototype with the new comparison system and figure 4 shows the window-shopper profile with the search function and the simplified explanation of how the service works.

	"STUDENTS"	"WINDOWSHOPPER"
Device/Channel	Google	Facebook
Mobile	• Options • Comparison	• Attention • Search function • Simple design
Desktop	• Users are older due to newsletter subscription. • Recommendation > Search function • Trustworthy	• **Attention** • Search function

Figure 2 DT Framework for the case study

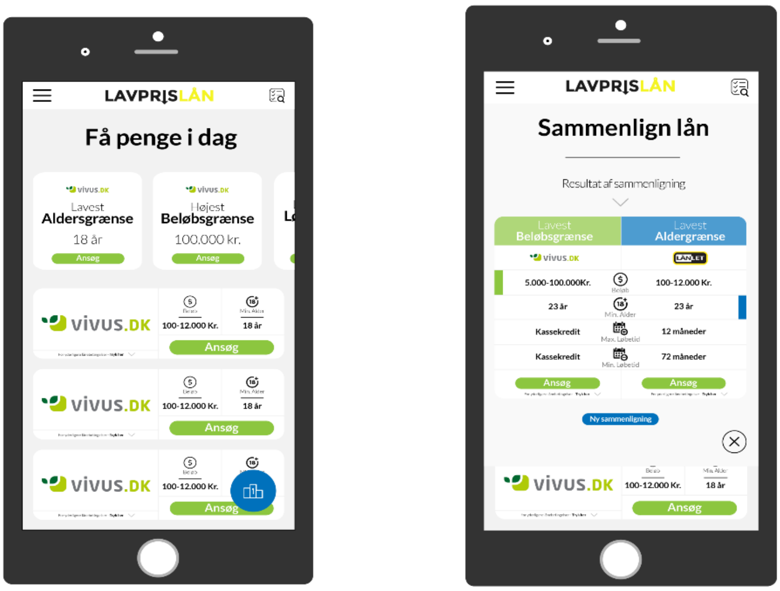


Figure 3 Prototype solution for the "students" profile and the comparison function.

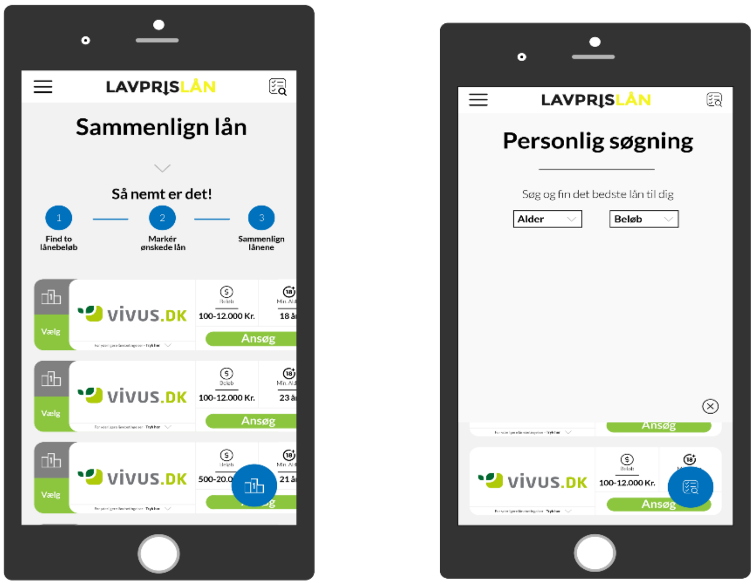


Figure 4 Prototype solution for the widow-shoppers profile; schematic search function.

Once the design prototypes were ready, the first cycle of the DT process concluded. The next step of the HITL method consisted on the inclusion of the DT results into the BD frame of the company to drive the new process of BD analysis. More specifically, we proposed the inclusion of the relationships learnt through the DT process into the second layer of the of the BD frame. This layer, as introduced before, contains the relationships learnt by the company overtime regarding the hypothetical fingerprint data of the users. As result, including this data produced websites whose design depended not just on the channel and device of the user's, but also on their clicking, scrolling and pointer behavior, as well as the rest of season-dependent relationships already known by the organization. This process opened new possibilities and questions for the organization, as for example design personalization depending on the country of access, the season of the year, or even cultural preferences. Additionally, the HITL approach also offered new opportunities for revenue optimization; for example, since products on the top of the websites had a higher probability of being purchased, the company could agree on a higher revenue percentage with the providers offered first.

Summarizing the results of the HITL approach, once the first iteration concluded, the company found improved answers in the sense of George *et al.* (2016), but was also confronted with new questions to explore. However, precisely those new opportunities and its requirements concluded with the failure of the implementation of the approach.

Exploration challenges and the "gravitational sink"

Analyzing the situation through George *et al.* (2016) perspective, this company was confronting new questions before the researchers arrival, as for example, how to make dynamic content websites. However, once the exploration of those new questions continued, the company decided to abandon the project due to 'lack of resources'. The organization was not ready for the technology's implementation due to two misalignments: First, the technology's capabilities were overestimated. The organization expected the DT process to be fully automatized. However, the current technology cannot provide unsupervised digital ethnography able to find relationships and stories on the data, produce a DT framework and embed this knowledge into solutions automatically. The scope of these problems belongs to the realm of ethnography and its digital variances. Second, the project required an increased amount of resources; the organization has six types of services using seven languages. The cultural nuances (granularity) of each service and language needed to be analyzed for each of the variables in the two layers of the company's data frame. This meant devoting several employees for the technology exploitation where new questions could potentially form needing even further resources. However, the company had assigned a few hours to the exploration phase of the project and none to its exploitation. Therefore, the scope of the problem and its granularity was not conveniently considered before its exploration. However, its potential to disrupt the way websites work in this industry attracted the company to its exploration. Analyzing the disruptive potential in the sense of Yu & Hang, (2010), the implementation of the tailor made websites would have had a lower performance at first; the new website designs offered a smaller number of products than the previous designs. However, through the loops of the DT process, the designs would have integrated the knowledge collected, translating the smaller number of offerings into a more accurate set considering the visitor’s preferences.

Taking these results into account, we propose in figure 5 an addition to George *et al.* (2016) framework; the gravitational sink. Here, we suggest that technologies exploring new questions (the unknown) and dealing at the same time with a high granularity and scope of data have a high probability of failure.

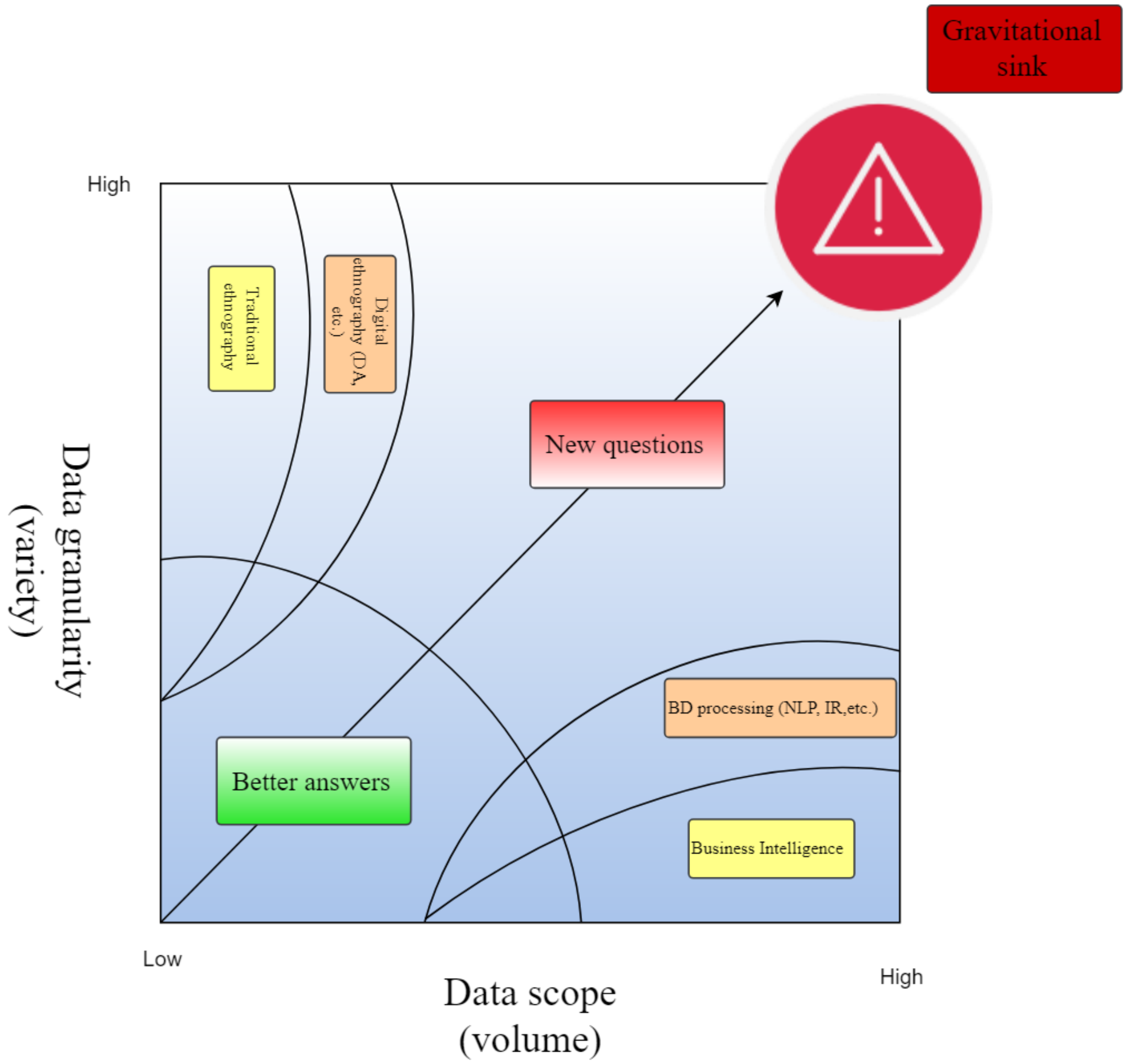


Figure 5 The gravitational sink for BD management

Managerial results: dealing with data scope and granularity

As figure 5 shows, when the scope of the data regards a small volume with no variety, it can be handled with statistics, or manual data processing. Once the data scope increases, BD processing can help handling large volume of data. Tools for business intelligence can effectively analyze this volume in absence of variety. Additionally, when the variety increases, tools as natural language processing or image recognition provide a solution. Once these technologies are in play, the resources allocated for data analysis grow due to the problems presented by Del Vecchio *et al.* (2018): need of skilled analysts, increasing data storage and processing cost, security risks, etc. When the data presents increased variety and a small volume, traditional ethnographic methods handle data thickness in the unstructured data (Wang 2016). Nevertheless, once the volume of data increases, digital ethnography is required to analyze larger datasets. Both sets of methods, digital ethnography and BD processing, help to begin the exploration of new questions, but none of them can fully give an answer to high volume and variety data related problems. These, require of an investment of increased resources not specified beforehand that presents high risk of failure due to the technology dynamism and continuous need of exploration. These methods, help explore new questions, but none of them fully answer high volume and variety data related problems. This exploration of the unknown represented by the gravitational sink presents high level of challenges, however, as the technology-driven literature promises, it also has the potential for long lasting disruption (Habtay, 2012).

6 Discussion

This case study has shown how the implementation of BD technologies represents challenges for the organization beyond technical regards. Using data with high scope and granularity implies the lack of standardized technologies to answer the new questions that BD brings. Therefore, each company must develop its personalized solution. However, the cost of such a solution is unknown, unforeseeable and goes beyond technological development costs or increased need of resources; it may entail a total shift from the company's original purpose. This reasoning is glimpsed in the current case, where, continuing to pursue the exploration of the experimental method also implied a radical change on the company's focus; shifting from an intermediary selling services, to a software development company producing the AI ethnographer. Therefore, the implicit dynamism of the BD technology produces continuous unforeseen changes and challenges in the organization. As the case has shown, if the company would have continued exploring the possibilities of the HITL approach, changes in the reorganization's core competences, its resources and even its business model would have happened. Nevertheless, the gravitational sink will continue attracting companies, as those technologies surviving the associated challenges, have the potential for harnessing increased market share (Habtay, 2012). In the present case, if the organization would have succeeded in implementing the tailor-made websites, they could have overrun the competition: Customers could have enjoyed an improved experience and the company could have increased its customer's knowledge. This way, selecting only those providers offering products preferred by the customers where the organization was operating or even opening to other markets. The tailor-made technological solutions dealing with increased data granularity and scope bring unexpected challenges, but also has the potential of shaking the market's expectations resulting on the re-evaluation of how that market works (Yu & Hang, 2010).

Additionally, we have hypothesized that ambidextrous organizations may enjoy a better position to explore the gravitational sink due to their experience dealing with the dynamism of exploring and exploiting simultaneously. However, as the case has shown, temporal ambidexterity fails to harness this advantage as exploration and exploitation challenges need to be commonly considered as a dualism and not in an intermittent manner. This way, the strategy considers holistically the challenges of the technology's implementation. Therefore, we propose further research in ambidextrous organizations following a structural or contextual strategy to devise whether they are in a privileged position to harness the disruptive potential of the BD technologies.

Limitations

We highlight problems with the size of the sample studied for the DT process, where only one iteration of the method was carried out and thus, its results are only applicable to this study's company. Therefore, in order to implement the HITL approach, we recommended the continuous reiteration of the DT process to use it as a leaning tool as proposed by Beckman and Barry (2007). However, here is used as an example of how the HITL experimental process could be carried out, showing that even with rapid prototyping, DT already offers visual improvements of the website designs. Additionally, in the current case, the new designs' performance is not tested against the old designs. Nevertheless, as the project was never fully implemented, this comparison was not possible. Finally, we would like to point problems with reproducibility of this research due to confidential agreements with the organization regarding the data used as well as the behavioral data collected.

References and Notes

Beckman, S. L. S., & Barry, M. (2007). Innovation as a Learning Process: embedding design thinking. *California Management Review*, *50*(1), 25–56. https://doi.org/10.2307/41166415

Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99–113. https://doi.org/10.1016/j.techfore.2017.07.027

Cao, Q., Gedajlovic, E., & Zhang, H. (2009). Unpacking Organizational Ambidexterity: Dimensions, Contingencies, and Synergistic Effects. *Organization Science*, *20*(4), 781–796. https://doi.org/10.1287/orsc.1090.0426

Del Vecchio, P., Di Minin, A., Petruzzelli, A. M., Panniello, U., & Pirri, S. (2018). Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges. *Creativity and Innovation Management*, *27*(1), 6–22. https://doi.org/10.1111/caim.12224

Feldman, M., Kenney, M., & Lissoni, F. (2015). The New Data Frontier: Special issue of Research Policy. *Research Policy*. https://doi.org/10.1016/j.respol.2015.02.007

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

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. https://doi.org/10.5465/amj.2016.4005

Habtay, S. R. (2012). A Firm-Level Analysis on the Relative Difference between Technology-Driven and Market-Driven Disruptive Business Model Innovations. *Creativity and Innovation Management*, *21*(3), 290–303. https://doi.org/10.1111/j.1467-8691.2012.00628.x

Horst, H., & Miller, D. (2012). *Digital Anthropology*. *Psychoanalytic review*. https://doi.org/10.1521/prev.2013.100.1.211

Jansen, J. J. P., Van Den Bosch, F. A. J., & Volberda, H. W. (2006). Exploratory Innovation, Exploitative Innovation, and Performance: Effects of Organizational Antecedents and Environmental Moderators. *Management Science*, *52*(11), 1661–1674. https://doi.org/10.1287/mnsc.1060.0576

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, *2*(2), 59–64. https://doi.org/10.1016/j.bdr.2015.01.006

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14.

https://doi.org/10.1007/s10708-013-9516-8

Kozinets V, R. (2013). *Robert V. Kozinets*. *Netnography: Redefined*.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*(2 S), 95–112. https://doi.org/10.1002/smj.4250141009

Mayer-Schonberger, V., & Cukier, K. (2013). Big data: A revolutionthat will change how we live, work and think. London: John Murray

McAfee, A., & Brynjolfsson, E. (2013). Big data: The management revolution. *Harvard Business Review*, (OctobeR 2012), 1–9.

McKinsey&Company. (2013). *Making data analytics work: Three key challenges*. *McKinsey Quarterly*.

O'Reilly, C. a &, & Tushman, M. L. (2004). The ambidextrous organisation. *Harvard Business Review*, *82*(4), 74–+. https://doi.org/Article

Park, G., Schwartz, H., & Eichstaedt, J. (2015). Automatic Personality Assessment Through Social Media Language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. https://doi.org/10.1037/pspp0000020

Plattner, H., Meinel, C., & Leifer, L. (2011). Design Thinking Understand – Improve – Apply. *Springer-Verlag Berlin Heidelberg*, *1*(April), 94. https://doi.org/10.1007/978-3-642-13757-0

Rothrock, Ling. Narayanan, S. (2011). *Human-in-the-Loop Simulations*. (S. (Eds. . Rothrock, Ling, Narayanan, Ed.). Springer. https://doi.org/10.1007/978-0-85729-883-6

Salganick, M. J. (2017). Bit by Bit: Social Research in the Digital Age, *Princeton*(Princeton University Press.), 1–12.

Somoza Sánchez, V. V., Giacalone, D., & Goduscheit, R. C. (2018). Digital anthropology as method for lead user identification from unstructured big data. *Creativity and Innovation Management*, *27*(1), 32–41. https://doi.org/10.1111/caim.12228

Turner, N., Swart, J., & Maylor, H. (2013). Mechanisms for managing ambidexterity: A review and research agenda. *International Journal of Management Reviews*, *15*(3), 317–332. https://doi.org/10.1111/j.1468-2370.2012.00343.x

Wang, T. (2016) 'Why Big Data Needs Thick Data' [electronic resource]. Available at: https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7 on the 27th of August 2018

Yu, D., & Hang, C. C. (2010). A Reflective Review of Disruptive Innovation Theory. *International Journal of Management Reviews*. https://doi.org/10.1111/j.1468-2370.2009.00272.x