**Headline: Big Data Literature review**

**Authors: Henrik Sølvbjerg Pahus and Lars Hoelgård Sunesen**

**Abstract: This shortened literature review is part of an ongoing research project into big data. More specifically, we are looking into generating an operational strategic model that companies within the hospitality sector can apply when working with big data as part of their business operations. The review initially defines the concept “Big Data”, and then subsequently dives into some of the prevailing literature available on the topic. Following the literature review, we discuss how the findings from the literature review fits with topic of our research project, and what might be learned from the previous research within this field.**

**Keywords: Big Data, Hospitality, Strategy, Consumer behavior, business development**

**Introduction**

Big Data is without a doubt a popular topic in the current technological debate. Together with AI, Big Data is one of those issues that has an increasing influence on how many companies and public institutions make their everyday decisions. Is has also drawn the attention of several scholars from within various professions, and new knowledge on the topic is constantly being produced.

This literature review is written as part of research project that spans from 2019 – 2023. The over all purpose of the research project is to formulate a model, which will enable companies to understand how to transform the massive amount of information that is to be gained from various Big Data sources, into something that functions on an operational strategic level. The output of this project will focus on generating solutions to companies from primarily the hospitality sector, on how to apply knowledge from Big Data sources into their pre-existing business model.

There are various approaches to creating a model that deals with the usage of information from Big Data. One might argue that since Big Data is primarily a technical topic, one ought to pursue a framework that puts focus on a technical model that centers on how to “crunch” the data more efficiently, or how to make algorithms that more accurately helps to calculate outcomes. Our focus however, will be on developing a strategic model that companies can use when working with Big Data as part of their business model.

Even though the focus area in this review is the hospitality sector, we are also trying to put a broader view on the concept of Big Data. This will yield a holistic understanding of current research within the area, and hopefully help us in our further research. Likewise, the model that we intend to develop would also be useful for a wide range of companies from without the hospitality sector. Additionally we found through our literature search process, that very little research concerning big data and strategy within the hospitality sector had been conducted. That led us to widen the search for literature that might help us in the further research process.

As the topic Big Data in an overall term, we will initially make a more narrow definition on the concept. This will be followed by the literature review that we have divided into three broad categories. Subsequently will follow a discussion on what we have gained from conducting this review, and how this might help in our current research project. Finally we will sum up the findings in a conclusion.

**Definition**

In terms of the concept of “Big Data”, a lot of research has been conducted during the past years, and Big Data is a wide concept that has been defined in various ways. However, an important notion is the similarities among different scholars interpretation of the concept. In his review from 2016, Martin Hilbert mention three characteristics, which describes the core of the Big Data concept. These characteristics are *Volume*, *Variety* and *Velocity* **(Hilbert, 2015, p. 3)**. Several scholars support Hilbert in this way of defining Big Data such as Miah **(Miah *et al.*, 2016, p. 771)**, and Kitchin and McArdle **(Kitchin and McArdle, 2016, p. 1)**. The below mentioned V’s are often accredited to Douglas B Laney.

*Volume*

The term “Big Data” is based on the understanding that a significant amount of data is being analyzed and interpreted. Hence, conclusions are based on a solid databased foundation. This viewpoint is supported by Chen, who refer to Big Data as *“datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse”* **(Chen *et al.*, 2016, p. 425)**. Vinod ads another aspect to this perspective in the sense that Big Data volume should be seen as dynamic. For instance, his study within the travel industry suggest that the amount of data increases in time, based on consumer actions and purchase decisions **(Vinod, 2016, p. 354)**.

*Variety*

One important aspect when working with Big data is the fact that data is much more than numbers. Hilbert mention data sources such as video, audio and text as notable sources for gathering data. Some support to this perspective is given by Miah, who states that *“Big Data is characterized by its Volume, Velocity, Variety and Volatility”* **(Miah *et al.*, 2016, p. 771)** and Chen who describes that Big Data *“…can be generated through multiple information technologies and systems…”* **(Chen *et al.*, 2016, p. 425)**.

*Velocity*

In addition to the perspectives of Volume and Variety, there is a shared understanding of the importance of Velocity as a characteristic of Big Data. According to Kitchin and McArdle, the term *Velocity* means that data is produced real-time **(Kitchin and McArdle, 2016, p. 1)**. A contribution to this viewpoint is one of the most commonly cited definitions by Gartner: *“Big Data is high-volume, high-velocity and high-variety assets that demand effective, innovative forms of information processing for enhanced insight and decision making”* **(Neely, 2016, p. 1384)**.

Among scholars there has been an addition to the definition and understanding of the Big Data concept. The fourth characteristic is Veracity, which is included in research conducted by Cheah**(Cheah, 2017, p. 230)** and Kitchin and McArdle **(Kitchin and McArdle, 2016, p. 1)**.

*Veracity*

There has been added more qualities to Big Data, and one of these is veracity. This is the understanding, that data can be messy and contain uncertainty **(Kitchin and McArdle, 2016, p. 1)**. Same perception of Big Data is mentioned in a study by Cheah. She argue: *“With availability of consumer data at high volume, velocity, variety and veracity, new business opportunities are presented”* **(Cheah, 2017, p. 230)**. Finally Schroeck states that, *“Often a fourth dimension is added to address uncertainty of the data, namely veracity, referring to the reliability of a certain data type”* **(Neely, 2016, p. 1384)**.

In addition to the four V´s listed above, resent research has is also been working with adding extra V’s. For now, we will however focus on the four classic V’s.

# **Literature study**

**Topics**

In the following, there will be an abbreviated review of the most relevant literature related to topics within the field of big data. As pointed out previously, this literature review is part of an ongoing research project into the strategic usage of big data in the hospitality sector, however we found it useful to view several aspects of Big Data from various industries, and not only from the hospitality sector. Additionally, it should be noted that this is by no means an exhaustive review of the literature on big data.

*Big Data and consumer behavior*

One important aspect that needs more attention, when speaking of Big Data, is the matter of behavioral statistics, which can be used by businesses to improve their performance.

* In a study conducted by Stoicescu **(Stoicescu, 2016)**, Big Data is analyzed by looking on the advantages and disadvantages of using it in terms of interpreting on consumer behavior. Even though there are similarities in terms of the current study, Stoicescu suggests a general approach. The results show a number of different aspects, which can be divided into previously mentioned characteristics Volume, Variety, Velocity and Veracity. This study examine the Big Data concept from a general perspective and suggest that businesses can improve consumer experiences by empowering new business models..
* Another contribution to this field of consumer behavior is the aspect of using Big Data to support the decision making process within e-commerce. This area has been investigated by Kannan and Raja **(Kannan and Raja, 2016)**. In their study, it is investigated whether case-based reasoning can be used to support consumers in the decision-making process. Findings suggests, among other things, that word of mouth marketing is an important source for store traffic.
* As a result of the increasing focus on Big Data among scholars and businesses, much attention has also been on different aspects of online behavior. Chen and colleagues **(Chen *et al*., 2018)** seeks to clarify which characteristics that have an impact on online consumer behavior. Their research find that many different aspects might influence consumer behavior. Some of their findings suggest that both website quality, consumer attitudes and innovation in Big Data all have an impact on consumer behavior.

Other scholars have conducted research based on specific technology such as mobile applications. In 2016, Chen **(Chen *et al*., 2016)** researched how to elicit intelligence and value from Big Data by analyzing customer behavior in mobile app usage. With a Big Data analytical approach, results showed that mobile app developers could learn about customer preferences and function usage. Another important outcome of this research is the difference in the way *Fans* and *Followers* behave when using mobile apps.

Another field within online consumer behavior is how consumers express their preferences. This field has been researched by Ma in 2019 **(Ma *et al.*, 2019)**. Research was conducted to clarify consumer preferences towards electric vehicles through online behavior.

*Big Data, hospitality and service*

Within the field of hospitality and service there has been an increased focus on the Big Data concept and how to benefit from large data amounts. Different scholars have made their contributions to this field.

* In Parks research from 2019 **(Park, 2019)**, it was investigated how revisitations to accomodations were related to previous experiences and satisfaction. Through extensive amounts of data, based on more than 100.000 online reviews, results showed that especially hedonic values and user burdens had notable effects on customer revisitations **(Park, 2019, p. 16)**
* An addition to the field of utilizing Big Data, to understand the relationship between hotel guest experience and satisfaction, is provided Xiang and colleagues **(Xiang *et al.*, 2015, p. 120)** . Their research seek to better understand hospitality issues by exploring and demonstrating the utility of Big Data analytics. Based on data from satisfaction ratings, most used keywords were extracted from customer reviews. Data showed that words such as “Room”, “Clean”, “Staff” and “Location” were some of the frequently used terms **(Xiang *et al.*, 2015, p. 126)**.
* Another perspective on Big Data is within personalization and customization of services and how to exploit Big Data through Customer Relationship Management (CRM). In a study by Anshari **(Anshari *et al.*, 2019)**, a literature review was conducted to shed light on this area. Key findings suggest Big Data is based on both structured and unstructured sources and businesses existing data from their CRM systems, However, the study also suggest CRM systems can be developed even further by mining data in appropriate manner. To be able to personalize and customize services, there is a need for customer profiling, which again require businesses ability to analyze and interpret large amounts of data **(Anshari *et al.*, 2019, p. 98)**.

In 2018 Ramona Marge **(Marge *et al.*, 2018)** and her colleagues came up with different trends that would arise as a result of the Big Data era **(Marge *et al.*, 2018, p. 125)**. Some of these trends can be seen as a natural next step for many businesses. First, data will reveal its secrets in the sense that different software makes it possible for everyone in the business to analyze on data. Therefore data is no longer just restricted to the interpretation of experts. Second, visualization tools such as dashboards will increase in popularity among businesses. These tools are intuitive and can be used to motivate staff. Third, usage of social media data to improve business performance will increase on a massive scale **(Marge *et al.*, 2018, p. 126)**.

*Big Data and business development*

Since the concept of Big Data was defined, there has been a lot of focus on different perspectives on which business development might be affected.

* Ralph Schroeder **(Schroeder *et al.*, 2016)** made an overall study with 28 participants from different businesses. He examined the current state of Big Data use in business, as well as opportunities and challenges presented by Big Data. One aspect that is highlighted is the general perspective of Big Data as being of low quality. There can be multiple reasons, such as lack of abilities to transform data into tangible assets or collecting the wrong data for specific purposes. Another issue is the internal policies that businesses execute. For instance how data is shared within the organization and how data is presented **(Schroeder *et al.*, 2016, p. 11)**. Another perspective to this field is, how to capture value from Big Data by start-up firms.
* Developing businesses through Big Data and decision-making processes by exploiting dynamic capabilities has been a field of study for Abbady, Akkaya and Sari **(Abbady** ***et al.*, 2019)**. They found that higher levels of dynamic capabilities are necessary for effectiveness of decision-making. In addition, they also found that firm type and age interact with dynamic capabilities and Big Data governance to predict greater decision-making effectiveness. Especially younger firms and firms with a large number of employees are more likely to use Big Data to make informed-decisions **(Abbady** ***et al.*, 2019, p. 437)**.
* Müller **(Müller** ***et al.*, 2016, p. 289)** suggest a set of guidelines for conducting Big Data analytics in information systems. Their findings support researchers in planning, executing and interpreting their own studies. By using 1.3 million online user reviews, Müller and fellow scholars by suggesting how to interpret different words in a certain order. This could be “bad buy make money”, which is interpreted into “negative effect” **(Müller** ***et al.*, 2016, p. 297).**

**Discussion**

As mentioned in the definition part of in this review, Big Data can be viewed through various prisms: *Volume*, *Variety*, *Velocity and Veracity.* All four are subject to constant changes and improvements in technology, and all four will have an impact when trying to build a comprehensive strategic model that tries to encompass the breath of Big Data. With the ever increasing development within technology, the ***volume*** of Big Data is most certainly going to increase, as the possibilities for collecting and analyzing the data will be enhanced. This will most certainly lead to more complexity, as the ***velocity*** (production of data in real-time) is also gaining pace. Combined with the ***veracity*** of Big Data sources, all three elements adds to the complexity of using Big Data in strategic setting. Even though we usually associate Big Data with hard numbers, and something that is analyzed in a statistical program, Big Data has the potential for much more. ***Variety*** is also an issue that adds to the complexity mentioned above, and currently in Europe, we have still not tapped into the possibility of for example using face recognition as a way of collecting information. This might off course be connected with the strict laws of the GDPR regulative imposed by the European Union. In other countries outside Europe however, that technology is widely used in combination with hardcore data – though still at the cost of adding more complexity.

The complexity highlighted above, is something that will become a pivotal part of our further research. As previously mentioned, our aim is to “simplify” the use of Big Data so that companies more easily can integrate it into their business model. Thus, each of the four defining characteristics mentioned above will have to be addressed and integrated into a model that might help companies.

Besides the different aspects of Big Data, we will also integrate some of the various approaches from the research articles mentioned throughout the literature review. However, one of the conclusions of this paper is that there is – so far – very little research into creating a holistic operational strategic model for hospitality companies. Thus, we will focus more on the methodological approach, more than on the findings in the articles.

**Further research**

Following our current literature review we can conclude, that very few scholars have focused on a model that enables companies to better use and understand Big Data outputs in a strategic setting. This is especially evident when looking at the current literature within the hospitality sector. This is something that we will focus on in our upcoming research project.

The research design is going to be an *explanatory sequential mixed-methods* design. Initially, we will do in-depth interviews with some of the experts within the field of Big Data. The interviews will be conducted with researchers, experts from the industry and the leaders who use and interpret the data. Secondly, we are going to conduct a questionnaire with some of the companies that are currently using Big Data as part of their business model. Thirdly, our aim is to formulate a cohesive strategic model that companies might apply when working with Big Data as part of their business model.