KNOWLEDGE MANAGEMENT WITH BIG DATA
CREATING NEW POSSIBILITIES FOR ORGANIZATIONS

Cecilia Fredriksson
Department of Political Science, Public Administration
Åbo Akademi University
cecilia.fredriksson(at)abo.fi

XXIV Nordiska kommunforskarkonferensen
Gothenburg, November 26–28th 2015

ABSTRACT

Scholars have not yet reached an agreement of a specific definition of Big data and the concept of Big data still raises many questions. The benefits of Big data, as well as the challenges associated with Big data, are known to some extent. Big data is seen as valuable for organizations, creating new possibilities and opportunities to develop knowledge management. Simultaneously, there are numerous challenges that should be addressed in order to create value from Big data. This paper offers a deeper understanding of the concept of Big data and how the application of Big data is reflected in the research so far.

Definitions and characteristics of Big data is presented in this paper. Economic, societal and decision-making aspects of Big data is discussed as well, to better understand the value of using Big data to develop knowledge (management) in organizations. Furthermore, challenges associated with Big data is presented, to broaden the awareness of obstacles that needs to be addressed when applying Big data.

KEYWORDS
Big data, concept of Big data, knowledge management

Draft – please do not cite or quote without permission from the author
1 INTRODUCTION

Big data have received a great deal of attention in recent years. At this point in time the amount of data is on a completely different level than before, due to the enormous distribution of information, the increasing amount of electronic devices leaving tracks and the overall development of our society. As stated by Vaidhyanathan and Bulock (2014, p. 57) “it’s a whole new set of tools for understanding our world”. The amount of information, in a different dimension than before, is said to lead to new possibilities and opportunities for actors and organizations in the public and private sector (Manyika et al., 2011; Mayer-Schönberger & Cukier, 2013). Large amount of information is nothing new per se. Today the data, to larger degree, is stored in a format which can easily be used. Thus a need of new approaches to collect, store, manage and gain value from all this data.

Even if the increased use of Big data is quite new, there are numerous indications that the use of Big data can create significant values for different organizations as it enhances productivity and competition for companies, while it strengthen efficiency and productivity for public organizations (Manyika et al., 2011). Scholars are repeatedly stating how the extensive amount of information can be beneficial for organizations (Davenport & Harris, 2007; Mayer-Schönberger & Cukier, 2013), and that Big data is considered to create new forms of value (Brown, 2014; Mayer-Schönberger & Cukier, 2013), even though distinct challenges exist regarding Big data management (Phillips-Wren & Hoskisson, 2015). Data is required to be managed in different steps and most of all analyzed (Kudyba, 2014), for organizations to gain knowledge.

The concept of Big data is seemed as abstract and even if the importance of Big data is generally recognized, there is no agreed upon definition of Big data so far (Chen, Mao, & Liu, 2014; Ward & Barker, 2013). Fosso Wamba, Akter, Edwards, Chopin, and Gnanzou (2015) agrees and states that despite the recent interest in Big data, little is known about what encompasses the concept of Big data. Furthermore, there are many misunderstandings about Big data and it is even stated that there are at least 43 different definitions (Huang et al., 2015). Big data is mainly used to describe enormous data sets (Chen et al., 2014), but according to Ohlhorst (2012), the concept of Big data has evolved to include not only the size of data sets but also the data management processes. Stated by Bi and Cochran (2014), the concept of Big data is about the characteristics of the datasets and the methodologies to process data.

With this paper, the author will clarify the concept of Big data. The lack of an agreed upon definition of Big data and understanding of the Big data concept, illustrate a need of a clarification of the concept. With the digitalization and expanding volume of data ahead, it is necessary to address the possible needs for new ways to manage data, production of knowledge and data security issues. This paper includes discussion of opinions of the Big data phenomenon and benefits and challenges associated with Big data, which the author believes will contribute to clarify the current absence of a comprehensive understanding of the Big data concept. Definitions and characteristics of Big data is presented in this paper. Economic, societal and decision-making aspects of Big data is discussed as well, as an attempt to understand the value of using Big data to develop knowledge management in organizations. Furthermore, challenges and issues connected to Big data is presented, to broaden the awareness of obstacles that needs to be addressed when applying Big data.

1.1 Content of paper

The literature for the review were selected based on the concept of Big data and topics related to Big data, e.g. case studies and focus from different fields, to better understand the concept of Big data on a general level. Relevant literature with both economic and societal focus has been included and the overall understanding of Big data is discussed. This is not an attempt to systematically review the literature on Big data, rather enable an open discussion and a broader
understanding of the concept. The Big data phenomenon existing today is fairly new, which is one explanation of the absence of core literature in this field of research, and publication years range from 2001–2015.

In the following chapter, various definitions and characteristics of Big data is presented. Chapter three present and discuss benefits of Big data with focus on economic and societal benefits, and benefits of applying Big data as support in decision-making. Chapter four highlights challenges of Big data, data management and the application of information produced by citizens. In the final chapter general findings is summarized and future research areas discussed.

2 WHAT IS BIG DATA?

Despite numerous discussions on the definition of Big data, Big data has no overall definition that’s agreed upon by the industry and academic scholars (Chen et al., 2014; Demchenko, Ngo, de Laat, Membrey, & Gordijenko, 2014; Mayer-Schönberger & Cukier, 2013). The definitions are many and multifaceted, and various attempts to define and operationalize Big data have been made (Kitchin, 2014a). More precisely, Big data is digital information from a variety of sources – social media accounts, web pages, purchases in stores, video and picture downloading, customer feedback, health records, maps, protocols and geographic locations in applications – just to name a few. Some of these sources have existed a long time and some sources are new as a result of the development and increase of social media, digitalization and new technologies. As an attempt to clarify this multifaceted concept of Big data, a variety of definitions and characteristics of Big data is presented.

Attempts to define Big data have included focus on everything from central characteristics and computing power to figures and numbers (Ward & Barker, 2013). According to Knapp (2013, p. 215), Big data refers to "tools, processes, and procedures that allow an organization to create, manipulate, and manage very large data sets and storage facilities". Big data is also described as an analytical tool by Jordan (2014). Big data is complex, as it needs data bases to be stored in, programs and tools to be managed, expertise and personnel being able to retrieve useful information and visualization to be understood. Big data as such is worthless.

According to Ohlhorst (2012), Big data is extremely large data sets, neither possible to manage nor analyze with traditional data processing tools. And the bigger the data set, the more difficult it is to gain any value from it. Also Provost and Fawcett (2013, p. 54) accentuate data sets and define Big data as "datasets that are too large for traditional data-processing systems and that therefore require new technologies". In the McKinsey report, Manyika et al. (2011, p. 1) refers to Big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze". And similarly stated by Phillips-Wren and Hoskisson (2015), Big data can generally be described as data that exceed or are beyond the capabilities of the organization to store or analyze for the purpose of decision-making. Phillips-Wren and Hoskisson (2015) refers to Laney (2001) making the first attempt to define Big data, describing Big data having volume, velocity and variety as primary characteristics.

boyd and Crawford (2012) define Big data as a cultural, technological and scholarly phenomenon, based on three assumptions. First, maximizing computation power and algorithmic accuracy to gather, analyze, link and compare large data sets. Second, drawing on large data sets to identify patterns in order to make economic, social, technical and legal claims and third, the widespread belief large datasets offers a higher form of intelligence and knowledge that can generate insights that were previously impossible. Big data is likewise described by Troester (2012, p. 2) as a phenomenon where traditional information management tools are not enough, or as "a
situation where the volume, velocity and variety of data exceed an organization’s storage or compute capacity for accurate and timely decision-making”.

Big data is not only a phenomenon of numbers and texts. Power (2014) describes Big data as the change in data management and according to Carter (2011, p. 5), Big data technologies refers to “a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high velocity capture, discovery and/or analysis” and “data sets whose volume, variety, velocity and complexity make it impossible for current databases and architectures to store and manage”.

As stated by Bi and Cochran (2014), there is many definitions of Big data, most of them referring to technologies to manage (capture, aggregate, process) the volume, velocity and variety of data. These characteristics and variations of them, is discussed in the following chapter.

2.1 Characteristics of Big data

Big data, and analytics associated with Big data, is by many scholars defined as a data source with three central characteristics; extremely large volume of data, extremely wide variety of data and extremely high velocity of data (e.g. Elgendy & Elragal, 2014; Hurwitz, Nugent, Halper, & Kaufman, 2013; Kitchin, 2014a; Laney, 2001; McAfee & Brynjolfsson, 2012; Phillips-Wren & Hoskisson, 2015; Raghupathi & Raghupathi, 2014a).

According to Chen et al. (2014), Big data has been defined as early as 2001 by the analyst Doug Laney (2001), who introduced the model of Three V’s. Laney called them the three dimensions in data management; data volume, data velocity and data variety. Laney does not describe the three dimensions (3D) as the Three V’s, but he is anyhow introducing the upcoming characteristics of the Three V’s. Although the model was not originally used to define Big data (Chen et al, 2014), Laney is clearly making a path for the definition of Big data and its characteristics in data management research. In 2001, Laney is stating that “[IT] organizations must compile various approaches” for dealing with data volume, velocity and variety. Furthermore, he highlights the economic struggle and difficulties to enable greater operational, analytical and collaborative consistencies because of the strained economic reality we live in. This is a reoccurring motivation for organizations in the public and private sector today, almost 15 years later, to apply Big data analysis even more, and use the existing data available for them. According to Mayer-Schönberger and Cukier (2013), the starting point of term Big data was shaped by science, such as astronomy and genomic, in the 2000s. Furthermore, they state that the concept of Big data ”is now migrating to all areas of human endeavor” (Mayer-Schönberger & Cukier, 2013, p. 6). Today the multifaceted concept of Big data is a part of most activity by organizations in the public and private sector, operating in a more electronic environment like never before. As a reason for this, users need to adress the volume, variety, velocity (also veracity and value) of fast growing data caused by the electronic environment, to develop knowledge management.

Volume describes the size of the data, which is a central feature of Big data today (compared to ten, or even five, years ago). Large volume of data simply refers to the large amount of data created every second. The second V stand for variety, indicating the various types of data (Chen et al., 2014; Hurwitz et al., 2013) such as structured, unstructured and semi-unstructured data (Chen et al., 2014; Ohlhorst, 2012) and the different ways to apply data. Velocity refers to data being rapidly gathered and processed in real time. Both Ward and Barker (2013) and Wang and Krishnan (2014) state that the Three V model, proposed by the information technology research and advisory company Gartner, is the most popular definition\(^1\) of Big data.

\(^1\) Wang & Krishnan (2014) are referring to the Three V model as a definition attributing to three fundamental features of Big data. Of a review of articles and research on Big data, many scholars use the Three V model as characteristics of
Some scholars also notice a fourth V; veracity (Elgendy & Elragal, 2014; Hurwitz et al., 2013; Ohlhorst, 2012) referring to how accurate and credible the data and analytics are. According to Chen et al. (2014) and Huang et al. (2015), the fourth V could also stand for value. In 2011, IDC (International Data Corporation), described by Chen et al. (2014, p. 173) as “one of the most influential leaders in Big data”, defined Big data technologies as “a new generation of technologies and architecture, designed to economically extract value from very large volumes of wide variety of data, by enabling the high-velocity capture, discovery, and/or analysis” (Gantz & Reinsel, 2011, p. 6). This definition is said to highlight the meaning and necessity of Big data, i.e. value (Chen et al., 2014). Huang et al. (2015) state that the Four V’s, volume, variety, velocity and value, are generally agreed upon, as a suitable description of Big data.

When reading literature on Big data, the Three V’s and the Four V’s are a re-occurring and common description of characteristics of Big data. These characteristics and variations explaining Big data, are confirming the fact that Big data is a multifaceted and complex phenomenon (Hurwitz et al., 2013). After all, one defining characteristics of Big data is its complexity (Desouza & Jacob, 2014; Troester, 2012).

Despite the range and differences in definitions, Ward and Barker (2013) have found some points of similarity in their attempt to shed light on the definition of Big data. One similarity is the size and the volume of data, which Troester (2012) on the other hand find less descriptive of Big data than e.g. variety, velocity, variability and complexity. Another similarity is complexity (also in Kshetri, 2014), referring to the complexity in structure and behavior of data, and also the need to correlate and share data across entities (Troester, 2012). A final similarity is the tools and techniques that are used in data management, i.e. technologies.

2.2 Unstructured, structured and semi-structured data

As aforementioned scholars agree upon, data are available in a variety of formats, such as structured data, e.g. transactional data and unstructured data, e.g. customer comments on social media and similar forums (Phillips-Wren & Hoskisson, 2015). Big data is to a large extent unstructured data. Because of its size, Big data imply complexity, and data is available in a variety of formats. The (most) common way to divide data is in structured and unstructured data (Ohlhorst, 2012; Phillips-Wren & Hoskisson, 2015; Salo, 2013).

“Structured data generally refers to data that has a defined length and format” (Hurwitz et al., 2013), such as data including numbers and dates, often stored in traditional databases and the easiest data to work with (Ohlhorst, 2012). About 20 % of existing data can be accounted as structured data (Hurwitz et al., 2013). Hurwitz et al. (2013) divides structured data in machine-generated structured data, which can include sensor data, web log data from web activities and point-of-sale data from product purchasing. The other part is human generated data, which can include input data such as information about you that you fill in online and click-stream data which is data generated every time you click a link on a website. According to Hurwitz et al. (2013), much of this kind of data can be useful for understanding patterns that have the potential of predicting outcomes, e.g. in decision-making.

Unstructured data, such as documents, videos and e-mails, on the other hand does not follow a specific format. Because the lack of technological solutions, it haven’t been possible to do much with the data, other than storing it or analyzing it manually (Hurwitz et al., 2013). Data need to be structured to allow us to work and use the information within data resources. Without structured data, otherwise useful information is useless. Machine-generated unstructured data

Big data, not so much as a definition. Nevertheless, there is no clear consensus on what can classified as definitions and as characteristics or features of Big data.
includes for example satellite images and pictures and videos and human-generated unstructured data includes social media data such as Facebook and Twitter, mobile data and website content. If 20% is structured data, about 80% can be considered as unstructured data, which is only growing (Brown, 2014; Gantz & Reinsel, 2010).

Some researchers use a third category as well, semi-structured data, which have a position between structured and unstructured data (Ohlhorst, 2012). Most of the data collected, stored, created and managed by organizations are more or less unstructured, making it more difficult to both retrieve and interpret (Beath, Becerra-Fernandez, Ross, & Short, 2012).

Alongside the concept of Big data, exist data mining. The concept of finding useful patterns in data is called data mining (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) or knowledge discovery in data bases, KDD2 (Han & Kamber, 2001). Other definitions of finding useful patterns are information discovery and knowledge extraction, among others (Fayyad et al., 1996). Data mining have evoked attention “due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge” (Han & Kamber, 2001, p. 1). Fayyad et al. (1996, p. 39) define data mining as “the application of specific algorithms for extracting patterns from data”. Data mining is one step in the cycle of Big data management, leading us to uncover interesting data patterns hidden in large data sets, therefore finding (new) knowledge.

2.3 What is Big data?

The literature offers a variety of attempts to define Big data, as seen above. However, many of them are describing characteristics, rather than clear definitions, as many scholars are referring to certain characteristics to explain Big data rather that agreeing on a specific definition. The lack of a general definition of Big data so far, can explain why a variety of characteristics of Big data are more common than a traditional definition itself. When the Big data phenomenon of today develop further, a somewhat agreed upon definition of Big data will occur, alongside a wider understanding of the concept of Big data. The Big data phenomenon of today haven't existed that long which can describe the lack of a clear concept so far.

The characteristics of Big data presented in literature so far are well-motivated and reasonable descriptions of what one can expect of Big data. The term “Big data” can be explicative, and insignificant at the same. As scholars suggest, there is much more to Big data than volume. This contributes to a deep complexity which motivate to further research of Big data and what it can bring, with curiosity but also cautiousness.

Scholars and experts (e.g. Chen et al., 2014; Mayer-Schönberger & Cukier, 2013) find no general definition of Big data so far. When reading literature on Big data and the complexity surrounding it, there are certain key elements often highlighted, mostly during the past five years, but also going back as far as early 2000’s. With aforementioned literature in mind, the author define Big data as “large amount of data, which can generate (new) valuable information and knowledge beneficial for organizations, compared to traditional data sets and information management”. The concept of Big data thereby encompasses the amount of data, the tools and technologies necessary to manage these large amounts of data and the value of knowledge, impossible to receive with traditional data management. Furthermore, the concept encompasses possibilities which did not exist before, but also challenges new to the phenomenon of Big data today. As the Big data phenomenon is rapidly changing as research evolve, it is likely that concept of Big data could also change in time. But as for right now, the focus is on finding ways for organizations and

---

2 According to Fayyad et al. (1996) data mining and knowledge discovery in data bases (KDD) have a somewhat different meaning. They refer to KDD as the overall process of discovering useful knowledge from data and data mining as a particular step in the process. Han and Kamber (2001) do not do this distinction.
actors in the society to gain as much value from Big data as possible. To enable this, it is necessary to understand the concept of Big data, why it exist, how it affects knowledge management, and the benefits it can create and the challenges organizations need to deal with.

3 BENEFITS OF BIG DATA

Big data is “widely equated with big value” (Brown, 2014, p. 213), even if challenges do exist finding the value because of the size of it (Ohlhorst, 2012). Baker, Kiewell, and Winkler (2014) agrees and states the follow; “for those able to bring order to big data’s complexity, the value is substantial”. Following chapter is divided into three parts, each presenting benefits that organizations can gain by using Big data. Economic and societal benefits are seen as central benefits in information management. As Balena, Bonifazi, and Mangialardi (2013, p. 528) states; “consistently with international and European Union’s objectives to deliver economic and social benefits, by harnessing the potential of information and communication technologies, research and development efforts are mushrooming worldwide to innovate governance and management practices”. Big data as decision support is also included as a benefit in this paper, partly overlapping both economic and societal benefits.

3.1 Economic benefits

The interest in data management is increasing as the amount of data is rapidly growing. The use of Big data has shown to contribute to efficiency for and development of authorities and an increased productivity and competition for companies. Organizations “are rushing to seize the opportunities offered by big data, and gain the most benefit and insight possible, consequently adopting big data analytics in order to unlock economic value” (Elgendy & Elragal, 2014, p. 222). There is no doubt that Big data imply economic benefits. A substantial part of the literature on Big data highlights the economic benefits with Big data, and how it can improve information and knowledge management in organizations. Big data can help organizations to be more strategic, enhance economic efficiency and optimizing recourses allocation (Bhimani & Willcocks, 2014; Kshetri, 2014). Information turns into valuable knowledge, patterns and insights and gathered information from Big data can be used to solve problems, increase profits and productivity, find opportunities and develop working patterns (Elgendy & Elragal, 2014; Ohlhorst, 2012) for both the private and public sector organizations. Data has become “a new form of economic value” (Mayer-Schönberger & Cukier, 2013, p. 5).

According to Chen et al. (2014), Big data mainly comes from and is also mainly applied in and by companies today. Companies have always been dealing with extensive amount of data in many different forms. “Big data is important because it enables organizations to gather, store, manage, and manipulate vast amounts data at the right speed, at the right time, to gain the right insights” (Hurwitz et al., 2013, p. 10). Digital technologies are a part of every aspect of business today (Bhimani & Willcocks, 2014), developing new possibilities for organizations to reframe their achievable economic efficiencies and make the decision-making more flexible. The application of Big data in organizations can enhance productivity, efficiency and competitiveness in many aspects (Chen et al., 2014; Manyika et al., 2011). In their report, Manyika et al. (2011) have identified five functions where Big data can be beneficial for retail, but as seen in other literature (Chen et al., 2014; Davenport, 2014; Elgendy & Elragal, 2014), some or all of these function can also be beneficial for organizations in general. The five functions are

1. Marketing
2. Merchandising
3. Operations
4. Supply chain
5. New business models
Within marketing, Big data can contribute with e.g. location bases marketing and customer segmentation. By using and comparing Big data, companies can predict consumer behavior and develop and improve marketing strategies and sales planning. Customer intelligence can be highly beneficial for organizations (Davenport, 2014; Elgendy & Elragal, 2014). Organizations can benefit from knowing their customers, what they need and what they think, to e.g. make better marketing decisions and predicting and reacting to upcoming trends. Big data can also benefit merchandizing, e.g. optimize prices and selection, and operation, by making the performance more transparent and optimize labor inputs. Big data can improve the supply chain, as it can optimize logistics and forecast stock. By using Big data in performance and quality management e.g. on supply chain, organizations can reduce costs, make the process more effective and improve productivity (Chen et al., 2014; Elgendy & Elragal, 2014). Finally, Big data can be beneficial in the sense of creating new and innovative business models. Manyika et al. (2011) mention price comparison services and web-bases marketplaces as central business models in the future.

There is evidence that Big data can play an important economic role for both companies and nations, enhancing productivity and competitiveness, beneficial for consumers as well (Manyika et al., 2011). The report by Manyika et al. (2011) suggests huge potential, e.g. reducing costs and increasing operational efficiency, for organizations in the society. Applying Big data can be beneficial in several different ways, as it can; (1) create transparency, (2) enable experimentation to discover needs, (3) segment populations to customize actions, (4) replace/support human decision-making with automated algorithms, and (5) innovate new business models, products and services (Fosso Wamba et al., 2015; Manyika et al., 2011).

The statement that Big data is a new form of economic value (Mayer-Schönberger & Cukier, 2013), seem to be approved by other scholars as well. A variety of economic benefits are enumerated, for both private and public sector organizations. Using Big data can create a positive increase of economic efficiency, efficient problem solving and an overall enhancement of data use and knowledge management, beneficial for organizations.

3.2 Societal benefits

Experts in the field of computer and information science state that Big data have potential to solve many challenges nations have in science, education, environment and medicine, with “enormous societal benefits” (Ohlhorst, 2012, pp. 80, 81). Big data is seen as a powerful tool offering the potential of new insights into a diversity of areas, such as cancer research, terrorism and climate change (boyd & Crawford, 2012). The digitalization the world is going through concern all areas of the society. More electronic information is produced and travelling across boundaries. Compared to economic benefits, which is more a relationship between knowledge and organization (company or nation), societal benefits also includes the citizen and the individual in a broader sense. There are two parts of societal benefits, those who govern, i.e. governments and other actors in the society, and those governed, i.e. citizens (Jordan, 2014).

In the following chapter, societal benefits are divided into four parts; Big data in the public sector, open data, smart communities and health care, giving examples of how Big data can be beneficial for the society.

3.2.1 Big data in the public sector

Even if the use of Big data seem to be more common in the private sector (Bassi, Leoni, Leucci, Pane, & Vaccari, 2014; Manyika et al., 2011) the opportunities for organizations in the public sector
sector are many, as they likewise are collecting and storing different sets of data. The data sets might differ from data sets in the private sector. Data sets such as electronic health care records and educational records are frequently increasing with the development of data technology. Scholars state that Big data can improve productivity and efficiency in public administration entities (Desouza & Jacob, 2014; Kshetri, 2014). The public sector has to deal with diminishing resources and economic strains, and the use of Big data and digital methods are seen as potential solutions for economic issues (Kennedy, Moss, Birchall, & Moshonas, 2015; Manyika et al., 2011). Not to mention democratic development as another benefit, as transparency and open information increases for citizens.

The public sector can receive many advantages using Big data, such as higher-quality services, increased public sector accountability, improved public trust in government and more well-informed citizens, policy makers and public sector executives. Despite the advanced e-government, “public sector agencies often do not make data available in an effective manner either across organizational silos or to citizen and business” (Manyika et al., 2011, p. 56) which is a problem. More accessible data can create more value for the public sector across departments and levels of government. Releasing data with as wide usage as possible, should be a priority (Manyika et al., 2011; Salo, 2013).

3.2.2 Open (government) data

One type of Big data is open (government) data. Chui, Manyika, and Van Kuiken (2014) are referring to what they call usable public data, produced by governments. Henninger (2013) and Bassi et al. (2014) are referring to public sector information, as information created, gathered, applied and owned by the government or public sector bodies. Public sector information or open government data⁴ has many purposes, not only for the citizens, but for the government itself. The government creates information by collecting data from citizens and institutions, which is later used for its own decision-making or for informing citizens and institutions (Henninger, 2013). One example of this type of open data produced by government authorities is the information from a smartphone application that uses real-time data provided by transit authorities to tell individuals at which time the next bus or train will arrive. According to Chui et al. (2014), open data have great potential in different areas and sectors, as such and combined with internal data, which is only applied within organizations and not for public use.

A fundamental principle for the public sector is transparency, openness and accountability and the main goal is to fulfill the interest and need of the citizens (Kierkegaard, 2009). According to Henninger (2013), one of the most important roles of public sector information is to serve the public interest and the main goal of open data is to enhance transparency and citizen participation (Bassi et al., 2014). By creating and sharing this information and data, it creates transparency and openness. However, a considerable problem is the lack of access to official data (Kő, Gábor, & Szabó, 2013), despite the fundamental principle of transparency in the public sector. Problems, such as inefficiency, non-transparency and not-citizen-driven, still exist (Kő et al., 2013). By permitting access to existing data (some data are naturally not accessible for everyone to see, e.g. health care records, due to privacy and data security factors) the public sector (as well as the private sector) can gain valuable knowledge, a more open decision-making processes and civic engagement (Chui et al., 2014; Kő et al., 2013). The revolution of open data is crucial for civic engagement, transparency and knowledge development (Balena et al., 2013).

---

⁴ Henninger (2013) don’t use public sector information and open government data as equal; she refers open government data as datasets of public sector information (p. 84) or public dataset produced by governments (p. 85). In this paper, the term open government data will be used when discussing public sector information, public data, government data or other similar descriptions.
3.2.3 Smart communities

With the development of information and communication technologies (ICT), cities have the possibility to become "smart cities". The smart city is composed by ICT, innovation and urban governance and a significant aspect is the production of data analytics to understand, monitor, regulate and plan the city (Kitchin, 2014b). Vanolo (2013) distinguish six characteristics – smart economy, smart mobility, smart governance, smart environment, smart living and smart people – describing the concept of a smart city. A smarter management of urban spaces in combination with ICT can be a positive development for urban management and urban governance, creating innovative smart cities (Balena et al., 2013).

The smart city development is interesting for both companies and public sector actors, as it enhance potential socio-economic development in the city, with new technologies and services by companies (Kitchin 2014). Smart cities are not only about the organizations, the city itself, a central actor in urban management is the citizen. A (open) data infrastructure enables citizens to benefit from the smart city, but also participate in creating the smart city, e.g. through internet-based mobile applications. Efficient use of e.g. geographical information is central in urban management, for improving urban quality of life and supporting community involvement (Balena et al., 2013; Kamel Boulos & Al-Shorbaji, 2014). An inspiring example of involving community members in the smart city development is the application developed by Balena et al. (2013). With their smartphone application, they are trying to show the importance between combining structured expert knowledge and public dispersed knowledge, to improve data infrastructures and services, and making it more transparent. This would, in the end, facilitate urban utilities management practices. They are testing a smartphone application with a bottom-up approach. Citizens using the application can give up-to-date recommendations and suggestion, e.g. by reporting on broken trash cans or streetlights around the city. Data is collected, functioning as a base for solving problems and improving the community in a fast way, what Balena et al. (2013) calls instant governance.

The fast increase of personal location data, which is produced by different devices such as mobile phones, can be beneficial for urban planning (Manyika et al., 2011). This data can give useful information for road construction and attempts to decrease traffic jams. Big data can contribute to a transformation in the knowledge and governance of cities, for a comprehensive understanding of urbanity (Kitchin, 2014b). Overall, the development of ICT and urban management can be significantly beneficial for the society, offering the possibility for community members to participate in the development of their own community. This relationship does not only benefit the community, but the citizens as well.

3.2.4 Health care

Health science is a common era where Big data have been widely applied, and an era with rapid development in data and knowledge management. Clinical Big data analyses are used to analyze the prevalence and trends of diseases among populations (Wang & Krishnan, 2014). New, and important, knowledge have been found as a result. Furthermore, clinical Big data can be used to determine causality, effect and association between risk factors and diseases, among many other examples of what Big data can bring to knowledge management in health science (Wang & Krishnan, 2014). Healthcare providers are reusing medical record data, originally collected for clinical purposes, to advance medical research (Beath et al., 2012). Big data have great potential to change the use of health science information and results, and therefore have a positive impact on health science (Huang et al., 2015).

By looking at what people were searching for on the internet, Google could predict the spread of the flue that started in the United States in 2009. With the large amount of data Google has, alongside the expertise and processing power, Google could provide public health care officials
with valuable real-time information. This is a typical example of Big data, and what can be done with it, to provide useful insights and significant value (Mayer-Schönberger & Cukier, 2013). The Google example is also an example of what can be achieved with Big data in the society, for instance for public health care services. Furthermore, integrating patient data across entities and analyzing health care data can improve the efficiency, quality and continuity of care, and predict outcomes (Elgendi & Elragal, 2014).

The health care sector is often the most expensive sector within a nation, with a large amount of employers, expenditures and an aging population. The health care sector is also a producer of enormous amount of data, such as health records and statistics, which have large operational efficiency effects, productivity benefits and potential to make the service better. Big data have significant implications for health care executives and policy makers (Manyika et al., 2011).

### 3.3 Support in decision-making

Big data as support in decision-making is integrated in both economic and societal aspects. Companies, governmental and other actors in the society, apply Big data as support in the decision-making in their organization. Since Big data as support in decision-making concern all organizations, and seen as beneficial for organizations, it is discussed as a separate benefit in this paper.

One area where the use of Big data has increased significantly is in decision-making (Power, 2014) and it is seen as having great potential improving decision-making in both public and private organizations (Desouza & Jacob, 2014). Both private and public organizations perform Big data analyses to collect and analyze information as material to support decisions. One reason for this is that Big data enables organizations to gather, store, manage, and manipulate vast amounts data at the right speed and time, which provides important insights in decision-making. The use of Big data analyses is therefore expected to have transformed decision-making processes in several aspects, but also to become more important in the future to create competitive and efficient organizations. Decision-makers want to use accessible data to make better decisions. Furthermore, there is an interest in finding new information and patterns leading to new insights and knowledge, valuable for organizations.

In general, ICT techniques are seen as support for more efficient and open policy making processes and innovative ICT tools can increase quality of participatory democracy (Kő et al., 2013). “Strengthening the government–citizens relationship is fundamental in order to establish the most suitable policy-making process” (Kő et al., 2013, p. 227). As a result, quality of policies can be improved, sharing of information and knowledge can be faster and citizens can be included even more. Kő et al. (2013) consider the popularity of social media and participatory tools, such as blogs, web/text/opinion mining, as a reason for decision-making actors in the public sector to take responsibility of achieving a change in society. “The emerging technological environment has dramatic impact on communication, information processing and knowledge sharing among public administration participants and also within civil society” (Kő et al., 2013, p. 228). Decision-makers in the public sector have access to large amount of data from and by citizens that should be used to improve citizen participation on government issues.

Big data will lead to better decision-making (Power, 2014). New data sources, technologies and analyses, will provide more and better decision support. Big data as support in decision-making can not only enhance more specific programs, but also public policy on a more general level. Large volume of data; collected, stored and managed, and most importantly analyzed (Kudyba, 2014), to develop knowledge functioning as a foundation for decision-making (Manyika et al., 2011; Raghupathi & Raghupathi, 2014b). According to Vaidhyanathan and Bulock (2014), we want to collect as much data as possible to make the best possible decisions.
According to Kudyba (2014), there is little value to Big data, unless data can help decision-makers to make better decisions. Big data should be used as support to make better decisions, develop strategies, enhance productivity and efficiency. However, there are some key elements (Kudyba, 2014) that have to exist in order for Big data to have a significant value:

1. Big data should contain relevant information corresponding to a particular process or activity. New sources of data doesn’t automatically offer descriptive information for every process or activity.
2. The data should also have high quality. Data sets with great volume or new sources of data and variables, must be a reliable and consistent support for decision-making.
3. Even if large amount of data offers knowledge not possible to find in more moderate and traditional data sets, drawing conclusions without careful analyses and any kind of expertise could have a negative effect on decision-making (Kudyba, 2014). To have any significance in the decision-making process, Big data must be managed properly with a variety of analytic approaches.
4. Hurwitz et al. (2013) recommends starting with a problem you want to be solved. It will help dictate the kind of data you'll need and how the working process should look like. Decision-makers, in whatever field, must determine what kind of question they would like to answer with information from Big data analyses (Kudyba, 2014). Kudyba (2014) define it as a conceptual model; there need to be a conception of which questions decision-makers would want to answer, in order to complete a desirable process of Big data management. And as stated by Provost and Fawcett (2013, p. 53), data-driven decision making refers to “the practice of basing decisions on the analysis of data rather than purely on intuition”.

Companies have for long understood the importance of the information they are collecting, and that the information companies are collecting can be used as support to make better decisions. Not all companies are taking this advantage. In Harvard Business Review 10/2012 (McAfee & Brynjolfsson, 2012) is stated that data driven companies are more productive and profitable than their competitors, which are less data driven. According to scholars, the evidence is obvious – to make decisions with the support of Big data, tends to lead to better decisions (McAfee & Brynjolfsson, 2012). Extensive amount of information is not used to the same extent in the public sector (Ohlhorst, 2012) as it is in the private sector, even if large amount of data is available, in health records, municipality records, social media and other forums used my authorities. The information is applied, but prerequisites to use it even more do exists.

3.4 Conclusions

As stated by Kő et al. (2013), there is still problems in making the public sector and the decision-making more open, due to inefficiency, lack of transparency and not including the citizens in the policy-making process. However, they state further, we are experiencing a dramatic transformation. The traditional data has become Big data, information has become open data, and knowledge is not only from a small and specific group of people. Everyone should be included in knowledge development nowadays. Furthermore, the policy-making process, regardless of sector, has developed from an intuition-driven process to a data driven-process. The change to a data-driven society, has not escaped anybody. While there are certain benefits with embracing the development of ICT and Big data, and using existing tools and data, there are certain challenges as well.

4 CHALLENGES

We live in a data-driven society created by us, however, the implications for society, individuals and businesses are not well understood (Power, 2014). Despite the common consideration that
the use of Big data is valuable, limitations and challenges do exist that we need to overcome to manage Big data properly and ethically (Jordan, 2014). Big data is confronted with many challenges and more research is necessary to improve the efficiency and management of Big data (Chen et al., 2014). In the following chapter, challenges regarding management and technology, quality of data, expertise and tools, culture and change, and ethical aspects are presented.

4.1 Management, technology and quality

That Big data is connected with technical challenges is clear (Ward & Barker, 2013). Because of the development of modern technologies for collection and managing data, data sets have grown in size, e.g. clinical data sets (Wang & Krishnan, 2014). Although considered valuable, Big data has many obstacles to overcome. A challenge is how to cope with the large volume of diverse data and time consuming data processing (Kernaghan, 2014; Kő et al., 2013; Wang & Krishnan, 2014). Wang and Krishnan (2014) bring up the important fact that many data sets only contain data of interest, i.e. secondary data sets. The original data sets contain raw Big data, too diverse to analyze and understand as such. Secondary data sets are extracted for analysis purposes, as a solution to the management challenges raw Big data brings. Other challenges connected to (clinical) Big data are sample size, selection bias, problem with interpreting data, missing values, dependence problems and methodologies handling data (Wang & Krishnan, 2014). When managing data, it is not only extensive, but often fragmented as well (Desouza & Jacob, 2014).

The case study done by Phillips-Wren and Hoskisson (2015) is identifying management challenges in creating a data-driven organization that has aspirations to incorporate Big data into decision-making. Management challenges are identified through interviews with elites in firms in the hospitality industry. With Chief Information Officers (CIOs) interviews, Desouza and Jacob (2014) found out that many CIOs consider poor data governance as a significant factor in limiting their efforts to pursue Big data. Data localized in different departments and collating this data is seen as challenging.

Technical challenges to leverage Big data and issues associated with data management (Bi & Cochran, 2014) need to be discussed further, since the volume, velocity and variety is only increasing. Traditional data management and analysis is not enough for the both structured and unstructured (and semi-structured) data that exist today. The emergence of Big data result in new challenges to traditional data management (Chen et al., 2014).

The quality of data is a challenge to deal with, to be able to make justified decisions based on the data. The less representative the data is the less valuable is it as well (Chen et al., 2014). Lai and Hsiao (2014) stress the importance of the quality of data. The data should have enough quality, appropriate for the purpose and for the decision-making. Because of the amount of data and therefore a complexity, it is important to be aware of the data and to clarify that they are comparable, when different datasets are merged together (McDermott & Turk, 2015).

4.2 Expertise and tools

A challenge associated with Big data is the need of expertise and new tools. When traditional methods to manage fast growing amount of information are not enough, problems regarding information management occur. A vast majority of companies agree that there is a lack of expertise and tools to gain value from Big data (Bi & Cochran, 2014). Earlier information management cannot be applied by organizations today, which justify organizations to develop new patterns and principle working with Big data. New tools and models are needed to manage the extensive amount of information. Researchers are arguing “a management revolution” (McAfee & Brynjolfsson, 2012, p. 62) has begun, and organizations are offered a set of new tools and opportunities.
There are tools available to deal with volume, variety and velocity of Big data (McAfee & Brynjolfsson, 2012), but there still developments that need to be made to handle Big data in a greater extent than before. Information goes to waste, as a result of lack of knowledge and expertise regarding the management of large amount of data. Talent to manage Big data is required, with the rapid growth of volume, velocity and variety (Kennedy et al., 2015; Kernaghan, 2014). The success of Big data (projects) depends partly on users’ expertise and the culture in the organization (Bi & Cochran, 2014). There will be a shortage of necessary talent to take advantage of Big data (Manyika et al., 2011). Expertise in statistics, machine learning, data management and analyzing is necessary to gain value from Big data. As stated by McDermott and Turk (2015), researchers in e.g. health services will need expertise (e.g. computer programming skills) to manage Big data.

4.3 Culture and change

Many organizations have not developed enough to manage challenges surrounding Big data and to embrace the Big data phenomenon, cultural and organizational changes need to be made. As mentioned by McAfee and Brynjolfsson (2012), the evidence is clear that data-driven decisions tend to be better decision. Still, many organizations are pretending to be more data-driven than they actually are. As stated by Beath et al. (2012), the lack of a collaborative culture can create obstacles to accessing valuable data. Organizations need to provide proper training for their employers, to be able to understand Big data and work with it (Fosso Wamba et al., 2015). A challenge is therefore to change the way of thinking in organizations and their activity, so they can benefit as much as possible from the possibilities and the knowledge Big data brings. It is necessary to question existing culture and working patterns within the organization for organizations to be able to deal with challenges such as expertise and tools needed to manage Big data and gain new knowledge.

4.4 Ethical aspects

Big data is described as something complex (Kshetri, 2014; Ward & Barker, 2013) and its complexity lead to possibilities, as well as challenges. These challenges give a reason to re-evaluate existing security models and tools (Demchenko et al., 2014). It is important for citizens to be aware of the fact that the information produced by them, are information used by organizations. Personal information by and about individuals, is used by organizations in different sectors. For natural reasons, some data requires special treatment, such as patient records, which are not intended to be available for everyone, even if it contains large amount of interesting and valuable information. The ethical aspects of Big data are central and important to take into consideration when extensive amount of information are processed and applied (boyd & Crawford, 2012) and privacy concerns are becoming more significant (McAfee & Brynjolfsson, 2012).

Privacy and IT has become a major part of the dialogue around the world (Kernaghan 2014) and an important theme in the Big data literature is the implications of Big data for individual privacy (Desouza & Jacob, 2014). This is referring to both protection of personal data when it is collected, but also when it is stored, transferred and used (Chen et al., 2014). When third parties are involved, it is important to take ethical aspects into consideration (Demchenko et al., 2014). The risk of finding a correlation between data and individuals, has led to concerns about privacy. Exemplified by Desouza and Jacob (2014), mapping gun owners in the community is not the best way to highlight open knowledge, as the publication of names and addresses of individuals owning a gun, could seriously harm them.

boyd and Crawford (2012) raises the concern that little is understood about the ethical implications underpinning the Big data phenomenon. With the fast growing volume of data, more
severe safety risks exist. Preventive measures should be taken to protect sensitive and personal data, to ensure safety for those who the information belongs to (Bi & Cochran, 2014; Chen et al., 2014). According to Jordan (2014), the important questions about Big data are likely to be ethical, not technical, and there are obvious risks to individuals’ data integrity and privacy. When discussing e.g. government data, it is important to be aware of how the government can ethically use Big data analytics tools to protect the public. Ethical knowledge is the key, given the strong role that data analytics will play in the public administration (Jordan, 2014). At the same time, there is a conflict between the request for secrecy and the request for more openness and transparency (Henninger, 2013). Big data is linked to privacy and security risks and the risks are likely to increase with the size, variety and complexity of data (Kshetri, 2014). Kshetri (2014, p. 1135) further state, that social and ethical issues are especially relevant “due to the underdeveloped regulations and regulatory infrastructure, which may give rise to consumer exploitation by businesses”. This may likewise be applied on the public sector organizations. It is worrying that survey results show that many (private) organizations are not prepared to address security and privacy issues (Kshetri, 2014).

In order to overcome Big data privacy challenges, leadership and transparency in organizations are important (Desouza & Jacob, 2014). As the amount of digitized data is increasing and moving within and between organizations, a profound evaluation, and necessary updates, of data policies in organizations is inevitable. Policy and privacy issues will become increasingly important (Manyika et al, 2011). Even if the information retrieved out of personal data, e.g. health records, can offer great benefits, the sensitiveness of that data can under no circumstances be underestimated. As this kind of data is in fact being digitized and rapidly moving across boundaries, ethical aspects of Big data is more important than ever.

5 CONCLUSIONS AND FUTURE RESEARCH

The concept of Big data is illustrated with complexity. The definitions and characteristics of Big data are many; Big data comprehend emphasis on data sets and the size of the data. Furthermore, it is described as a tool, process and procedure to manage large amount of data. Scholars also emphasize Big data as a phenomenon and as a change is the management of data.

The interest in Big data has increased over the years. Even if the large amount of data is increasing rapidly, many organizations are still unaware of the possibilities associated with Big data. The concept of Big data is fairly new and organizations are only starting to understand the value Big data can bring to knowledge management. And likewise important, organizations should be aware of what they need to accomplish regarding changes in culture, expertise, data security and management tools, to even enable Big data management. It is necessary to understand the concept of Big data, why it exist, how it affects knowledge management, and the benefits it can create and the challenges organizations need to deal with.

Desouza and Jacob (2014) argue that the perception that the use of Big data is valuable, actors in the society, such as policy makers and citizens, have started to consider the ways in which Big data can be used to improve public sector outcomes, such as public policies. This gives motivation to study the actual effects of Big data applications. As scholars are trying to clarify the concept of Big data, and frequently considering Big data as potential solutions to public and private issues, little is still known about the outcome of the practical use of Big data in different organizations. Most scholars agree on the complexity of Big data and what it can lead to. Complexity means value, as it gives us possibilities to manage and apply Big data as never before. On the other hand, complexity also signifies a variety of challenges and limitations, as mentioned above. As for today, Big data can change society, the way of thinking and making decisions in companies and public authorities when managed properly.
Big data seem to be more applied in the private sector in comparison with the public sector. Manyika et al. (2011) highlights the fact that companies are already using Big data and others need to follow, such as governments which also have significant opportunities using Big data. Furthermore, as stated by Bassi et al. (2014, p. 41) "governments around the world are starting to recognize the value of the data kept in public institutions". The unbalance should be evened out, so that public authorities could benefit from the available information. However, a wide lack of expertise regarding Big data and how it can be used in favor for the organizations do exists. Organizations should include Big data analyses as support in decision-making in a wider extent as before. To do so, a change in the decision-making culture is needed. Furthermore, an effort on the tools required to manage the data, needs to be made.

One significant concern associated with Big data is ethical aspects and how organizations can protect individual data and personal information. As Kernaghan (2014) states, privacy and IT has become a major part of the dialogue around the world, but the discussion should be even more connected to the public sector. Ethical aspects are a matter for any organization, both in the private and public sector. Companies should also be aware of data security issues, as they are using personal information to the same extent as other organizations dealing with clients and customers. Nevertheless, there is a need to combine the discussion of value and ethics and public sector, with the increased use of and request for open data as one reason. Ethical aspects should of course be taken into consideration when dealing with any kind of data, not only open data. More research is needed to better understand the impact the developments in IT has on ethics in the public sector and Big data security (Demchenko et al., 2014; Kernaghan, 2014).

As the application of Big data is at its starting point, further research on the use of Big data in both public and private sector organizations is needed (Chen et al., 2014). There is a request for scholars to look closely at how data are currently being used and assess the degree to which it is being underutilized (Desouza & Jacob, 2014). And as stated by Fosso Wamba et al. (2015, pp. 234,235), “very few empirical studies have been conducted to assess the real potential of ‘big data’”.

We need new ways to make data more structured, new technologies and methods of managing data and protecting fast growing sensitive data (Gantz & Reinsel, 2010). Finding value from Big data are unfortunately seen as the exception rather than the rule (Beath et al., 2012). A clear understanding of the concept of Big data, and benefits and challenges associated with it, could hopefully contribute to a development in this area. New possibilities and opportunities arise for organizations as the amount of data increases and organizations develop. Organizations have every reason to evaluate their current data management, to improve efficient data use and develop knowledge management.
REFERENCES


