Kevin Blasiak

Big Data; A Management Revolution

The emerging role of big data in businesses

Helsinki Metropolia University of Applied Sciences

Bachelor of Business Administration

International Business & Logistics

Thesis

06.05.2014

Helsinki Metropolia University of Applied Sciences

Author(s) Title Number of Pages Date	Kevin Blasiak Big Data; A Management Revolution 44 pages + 2 appendices 06 May 2014
Degree	Bachelor of Business Administration (BBA)
Degree Program	International Business & Logistics
Specialization option	International Business & Logistics
Instructor(s)	Kaija Haapasalo, Senior Lecturer

Big data is a term that was coined in 2012 and has since then emerged to one of the top trends in business and technology. Big data is an agglomeration of different technologies resulting in data processing capabilities that have been unreached before. Big data is generally characterized by 4 factors. Volume, velocity and variety. These three factors distinct it from the traditional data use. The possibilities to utilize this technology are vast. Big data technology has touch points in different businesses across industries, but finds its place likewise in government organizations and the healthcare sector. The development of sophisticated big data tools which change the corporate culture in organizations and will have a significant effect on the managerial decision making in businesses.

The targets of this research was to identify challenges in the field of big data, general technical and business related issues, to elaborate on case studies and provide a first guideline to the implementation of big data into a business organization. A model of data in the decision making process furthermore points out the importance of company internal factors to the implementation of big data.

The research considered secondary research from professionals in the field of big data and is completed by primary research in the form of interviews with big data professionals from the IT industry.

The main findings are the influence of corporate culture on the decision making process when utilizing data as management decision support. Furthermore, the research identified social challenges for big data and challenges regarding work ethics. Big data creates a lot of new opportunities for businesses, from the creation of business models to more efficient operations. Case studies from various well-known companies demonstrate the rich possibilities that big data provides and serve as guide for big data technology implementation in different SME's in undefined industries.

Keywords	big data, management, data science, information management, decision making, change management, corporate culture, analytics, privacy, decision making process

Contents

Appendices

List of tables and figures

1 Introduction

The term big data was coined in 2012 and has since become one of the most trending topics in technology, business and management. As demonstrated by the sudden incline in Figure 1, Google trends reveals the emerging interest in Big Data since 2012.

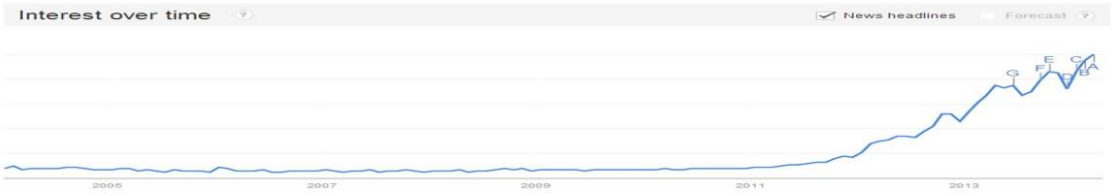

Figure 1. Google trends search term: Big Data (Google, 2014)

This interest for big data and associated topics is a result of the significance big data is believed to have on our society in the near future. Supposedly big data will change the way people live their everyday lives, their work and the way their homes are connected. Probably one of the most discussed big data technologies in the civil world is Google's self-driving car[1], which is an excellent example of how big data technology might change the way we used to do every day activities, like driving a car. Like the computer, there are indicators that big data will shape our lives and the business world like the first personal computers did at the end of the 20th century (Brynjolfsson, 2013a).

While this change will require some degree of adaption to this technology in its usage and the business culture, there are also serious threats to the individual. Comparing big data technology to the internet and its technological capabilities, the internet challenges lawmakers and industries alike. Popular topics in the media are copyright laws and entire industries losing profits to the black market and the online trading of copyrighted material (Institute for Policy Innovation, 2007). Big data, as revealed by the NSA affair in 2013 (Peterson, 2014), has the potential, if not regulated by governments and the society, to create a conflict of interest between the economy and the individuals. The discussion is similar to the current topic of copyright laws and the internet. Big data can give detailed insights into an individual's life and habits to external parties. Big data holds an immense value and therefore requires advanced security measurements to protect this sensitive data.

[1] A Toyota Prius equipped with sensors and data technology that is able to steer autonomously

Through big data we will be able to predict the future more accurately than ever before. Companies like Royal Dutch Shell, which have for many years predicted future developments in the society and economy, will be able to give more accurate outlooks with the help of big data (2013). Naturally, companies' interest in big data is very high, as it will enable companies to accurately predict the near future, monitor trends in real time and acquire substantial insights into their own business environment and the of competitors.

Companies that naturally, due to their business model, utilize data are the innovators working with this technology. Well-known companies like Amazon, Google, Teradata and large conglomerate corporations like GE have not only invested heavily into big data, but are already using it in their operations (Gartner, 2013). It is very likely that this development will spread to other more traditional business sectors. Especially in Logistics, E-Commerce, Insurances and Healthcare, big data has already started playing an important role and is transforming businesses.

According to an Accenture research (2014), even traditionally less innovative organizational bodies like government owned businesses and funds, in this case the US pension fund, start to benefit from the developments in big data. For instance, Germany's federal labor agency was able to cut costs significantly by applying big data, as show in figure 2. Big data has cut costs sporadically within a very short time.

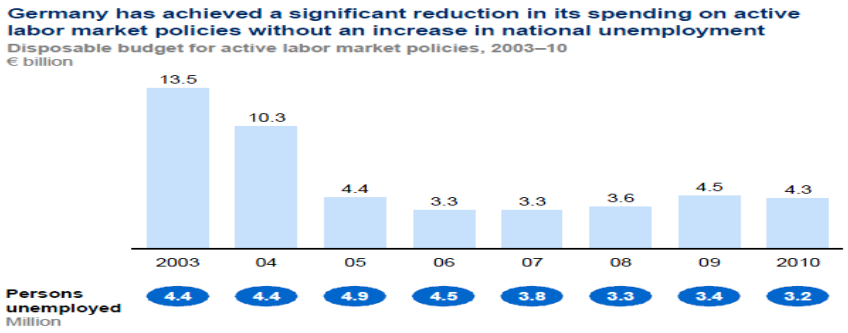

Figure 2. German "BA[2]" budget savings (McKinsey Global Institute, 2013)

If we talk about big data technology, the focus is not on a single piece of software or device. Big data technology is rather the accumulation of different technologies, together

[2] BA: Bundesagentur fuer Arbeit (Federal Employment Agency)

enabling the collection, storage and analysis of large sets of data. Most components of the big data technology have been available for some time already, however in the year 2012 a level of development was reached that could be described as big data technology leaving the prototype stage and actually having significant advantages to traditional data collection methods. The growth rate of big data is not linear, but exponential, thus it is growing at an increasing rate (Brynjolfsson, 2012).

Erik Brynjolfsson of the MIT Sloan School of Management aligns the significance of the development in big data technology with other great inventions like the steam engine, electricity and the computer that have shaped the development of our society like few other inventions (Brynjolfsson, 2013a). Keeping this perspective, big data technology is an increase in productivity. It enables people and organizations to work at a higher rate of efficiency and thus is so important for the economy and ultimately the society.

1.1 Objectives, scope and limitations of this research

The research in big data is justified by the fact that big data, especially in management, just very recently entered the spotlight in business. The general awareness of big data is increasing, however big data is only at an early stage of its development, with only limited research available, especially in the implementation of big data in the managerial decision making process of companies. This research is furthermore justified by the significance big data is believed to have to the business world and the opportunities it creates not only in business, but in the society. The aim of the research is to provide a general overview of the capabilities big data has and the opportunities that derive from its implementation. Creating an understanding of the role big data is believed to have in top management decisions and determining its position within a model of decision making.

The objectives of the research at hand are based on interviews, conducted with subject matter experts, which will be specified in the literature part of this thesis, as well as published work in the field of big data and case companies from different sectors of the economy. This thesis will take a closer look at big data and its relevance to business. We will discover: 1) the basics of big data technology and its merits to a business environment, as well as challenges of implementing big data into the business and the society in general. In particular, challenges in the fields of organizational culture, laws and regulations as well as privacy and security concerns are of interest.

Through interviews and articles of professionals in the field we will elaborate on 2) the public as well as professional opinion of individuals working in the sector and take a closer look at innovators, mostly large corporations that have successfully implemented or plan to implement big data into its strategy and business model. Analyzing the learnings of these organization, we will be able to draw a connection to a second generation of companies, the early adopters, which will further grow with big data. The second tier of companies may already include SME's, with smaller capital investments into big data research. The cases, together with the theoretical information can serve us as guideline and recommendation for the implementation of big data in different, big or small organizations in an unlimited variety of sectors. Looking at the cases in detail, possible big data touch points within an organization. This enables us to evaluate 3) the significance of big data for businesses and industries and find out whether big data has a future in business and ultimately management. To determine the significance big data actually has on business performance and management, large scale studies would have to be conducted. However, the case examples featured in this thesis demonstrate the advantages companies with a big data strategy have in comparison their competitors. The significance of big data is determined by the amount of new possibilities that derive from big data, mostly by giving companies better information on their business environment. Particularly interesting is how big data may represent a revolution in management as well as the decision making process in an organization. In the following pages we will also take a look at the practical changes triggered by 4) big data in the decision making process in the managerial departments of organizations. Concluding this paper with the development of a management model that connects big data to the theory of the decision making process, demonstrating the new role of data in organizations and its interconnection to different elements of the organizational structure defining big data.

This research is defined by the following four elements: 1) Introduction to big data, defining big data, as well as an introduction to technologies and software used in the process of collecting, analyzing and visualizing data. 2) Opportunities for organizations created by big data and example case studies of successful implementation of big data and the advantages it created for the particular organization. Followed by 3) outlined challenges of big data and critical evaluation, based on interviews with subject matter experts. Lastly, as a conclusion, this paper will outline the (future) role of big data in management and extend the model of managerial decision making by the element of data. Therefore the scope of this thesis can be described as a general introduction into the

field of big data, from a business perspective, covering the influence of big data on management decisions, business culture, legal and ethical environment as well as business model generation.

Based on the fact that merely a very limited amount of interviews with experts in the field of big data was conducted, the portrayed opinion might not represent the general opinion of professionals in the field. The opinions stated by the experts, represent only their own understanding of big data in their respective fields of expertise. If the aim was to gain a more comprehensive insight into big data practices in organizations today, a significantly larger number of interviews would have to be conducted, across industries and countries.

1.2 Research methodology

The research method applied in the paper at hand can be described as qualitative, including secondary research based on literature and online resources. The preliminary research, utilizing various online sources and published work in print is the foundation of the research in big data, identifying the current stage of development in big data technology, as well as apprehensions related to the technology. Case studies of various very different organizations from different sectors of the economy revealed the potential of big data technology and the opportunities that arise from a successful implementation of big data analytics into the business. A comparison between the approaches of leaders in different industries demonstrated a clear picture of the variety of big data solutions that enabled organizations to improve their current business model or identify new business opportunities.

An investigation into different software solutions offered by the leading business software developers gives the reader practical insights into big data analytics at a company level and moreover helps to identify crucial aspects of big data and its touch points with managerial staff.

In order to understand the role of big data in the decision making process of managers, firstly a study of the decision making process was conducted, with a focus on decisions in management, enabling us to draw conclusions between big data and the decision making process and ultimately extending the model of decision making with the element of data. This led to a point where the research could reveal the significance of big data

in the decision making process and relate it to the overall importance of big data in business.

Primary research, conducting unstructured interviews with experts from the field of big data provided more practical insights into the utilization of big data in companies that could be considered representative examples of their respective industries. The interviewed individuals and their companies are employed by a market leading German telecommunications company, which does not only maintains domestic operations in Germany, but also has extensive international reach and subsidiaries around the world. Naturally an industry with a high rate of data collection, the company's services range from telecommunications infrastructure services to services for private customers like stationary broadband and mobile broadband services. The second interviewed individual is self-employed and currently a big data advisor to a number of German SMEs.

Considering the findings of the primary research and combining the gained insights with the theoretical knowledge of the secondary research this paper concludes the research with suggestions and thought-provoking impulses to a theoretical future implementation of big data into a business/society and gives a brief outlook into the capabilities of big data and the opportunities that possibly arise from an implementation.

Finally, primary research was conducted in form of interviews to investigate the real effect of big data on organizations not being among the innovative leaders in the industry. Most cases presented by the literature are organizations that can be considered leading in their respective fields, however by conducting interviews with organizations that claim a more conservative market share, the research also considers the level of development in more traditional organizations. This way we aim to get a more general understanding of the stage of development in big data technology.

The secondary research, including journal articles, published books, video recordings of conferences and presentations, online resources such as blog entries of professionals in business, whitepapers and the like are mainly limited due to the fact that they were collected for different purposes and may represent only the opinion of a small proportion within the group of big data professionals. The research may also be limited by the fact that extensive research is still a scarcity in the new field of big data analytics. Therefore most research's origins are within the top research and educational institutions in the United States, such as MIT Sloan School of Management and Harvard Business School.

Purposely, this research is limited to the extent of detail in technical matters. The technological aspect of big data will only be covered to a very basic extent in order to provide a very basic understanding of the technological aspects of big data, which in the progress of this research will provide a basis to focus on the managerial aspects of big data.

1.3 Literature review

The research at hand is based on literature from various sources. The research features books from well-known opinion leaders in the field of technology and big data, with resident status at leading institutions such as MIT Sloan School of Management. Furthermore, the literature used for the purpose of this research include scientific journal articles, blog entries of established authors in the field of big data and various online resources maintained by well-known publishers.

This research is to a great extent guided by the work of Andrew McAfee and Eric Brynjolfsson, which both are associated with the MIT. They have collaborated on a number of works in the field of big data and technology in general. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies (2014) and Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy (2012) are two works by the before mentioned authors that give a good insight into the role that technology plays in our society and business. Both publications see the big data evolution and the evolution of technology in general as driver of our society's development. Better technologies are mainly the reason for increased productivity and always changing environments.

Eric Brynjolfsson has also great significance for this research as he was the first to study the correlation between big data as decision support and better business performance. His first study on this subject unveiled that businesses with a big data strategy are usually better performing than their competitors that have not (yet) implemented big data in the decision making process (2011).

Kenneth Cukier and Viktor Mayer-Schoenberger have very recently published literature with a focus on big data. In "Big Data: A Revolution That Will Transform How We Live, Work, and Think" (2013) they present a very comprehensive description of big data technology and the changes it will trigger in our lives and work environments. The work has

a very optimistic perspective on big data, however also draws some attention to the challenges that derive from big data technology.

Most of these authors and researchers also maintain professional blogs hosted by publishers such as Harvard Business Review and Harvard Business Review Blog. In journal articles and blogs, big data is discussed extensively and often from a very down to earth and practical perspective that focuses on big data in the workplace and how to cope with this new technology from a manager's perspective. Especially the published work by Harvard Business Review gives us a good understanding of latest big data trends from a business perspective without in depth technology analysis and IT specific information

For the purpose of this research, most HBR articles are very recent and have been published in October 2013 under the title "Big Data: The Management Revolution". The Big Data Journal, which is entirely dedicated to this new business trend, gives more specific insights into the role big data will play in organizations and is presenting findings from a data science perspective. This rather new discipline, which is most likely the area that will host big data as an academic field, points out the benefits big data can have on business decisions. It however also points out that big data is only at a very early stage of its development. The authors demand that big data has to find a way and establish itself as academic discipline, which will help organizations and schools to provide the skills that big data will require in order to be used efficiently. This is in accordance with an article in the Ivey Business Journal with the title of "Why big data is the new competitive advantage (2012).

Since big data is a topic with high relevance to business, but the core development happens in IT, a general understanding of the technology behind big data is beneficial to the reader. Jeffrey Needham has a series of books with a big data focus. "Disruptive Possibilities: How Big Data Changes Everything (2013)", outlines the basis of the technological issues related to big data and provides fundamentals to understand big data as a technology.

To get a better understanding what analytics software and business solution offer, this research also contains white papers and online resources provided by the leading organizations in the big data software applications industry. Especially IBM and SAP have an extensive offer of whitepapers that explain the implementation of big data into the business and its merits for the implementing organization.

.In general, most published works and opinion leaders have a very positive attitude towards big data and its development. Neil Richards and Jonathan King however approached big data from neither a technological nor a business perspective. In a Stanford Law Review paper with the title ”Three Paradoxes of Big Data” (2013) both voiced a critical opinion towards big data and its impact on privacy and security. The benefits for the business may be tremendous, but big data has also the capabilities of being one of the biggest threats to privacy in recent years. The paper suggest that the society has to develop a legal framework and start an ethical discussion about the use of data and how it can interfere with the interests of the individual.

2 Fundamentals of Big Data

To gain a better understanding of big data technology and how it is defined, this part of the research focuses on few of the key technologies that made big data possible and provides a definition of big data as well as an introduction into the field of big data.

2.1 Defining big data

Big data is a rather new term, as indicated by Google trends (2014), however it is still surprising that up until now there is no clear or uniform definition of big data. Opinion leaders and companies working in the field have their own opinion and definitions of big data at this point. Clear is however that big data embodies an accumulation of different technologies under the heading “Big Data”.

Especially the term “big” creates problems. The main reason for that is simply the rate of growth of data. Data volume that is considered to be big data today, will be comparably small data in the very near future. Therefore, “big” does not describe the real size of data but its relative size to the capabilities of technology of the present day. Big data represents a challenge in the technological capabilities to store and analyze data. Most opinion leaders are however in accordance with the 3 V’s of big data: Volume, Velocity and Variety, which was coined already in 2001 by Gartner analyst Doug Laney. Therefore, in most cases big data is defined through these 3 V’s of big data.

2.1.1 The three V of big data

As show in figure 3, big data is usually determined by 3 equally important factors. The combination of the 3 is what makes big data different from the traditional data.

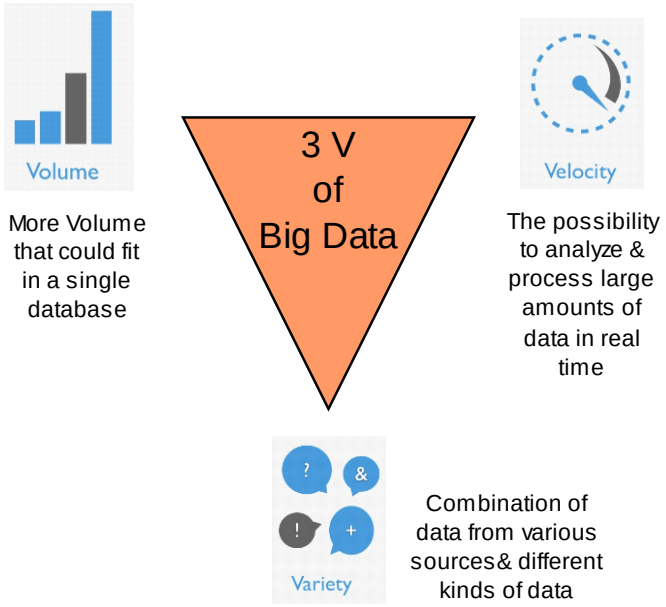

Figure 3. The 3 V of Big Data

Volume:

Volume simply describes the sired volume of big data and the possibility to store and analyze larger sets of data than ever before. Through volume a better and more detailed insight into an issue is created. We can achieve a higher accuracy of information gained from data and thus are able to understand schemes a lot easier.

Velocity:

Velocity describes the technological capabilities to collect and analyze data in real time. As mentioned before, volume is an element of big data, however technology nowadays also enables us to get insights into data in real time. Organizations used to rely on historic data of previous months, etc., simply because technology would not allow to process data fast enough. With the new technology we are able to analyze data in real time (Gartner, 2013). This is especially crucial in time sensitive businesses like the banking industry, where markets can change in a matter of seconds, but also the logistics sector is benefiting from real time updates and information flow.

Variety:

Probably the most crucial element of big data is variety. Variety describes in general the diversity of data. The multitude of sources of data, giving us the possibility to interconnect different data sets and derive conclusions from it. In previous decades, organizations were already collecting large data sets from various sources like mobile phones, emails, etc. Technology now enables us to actually take a look at the data and combine it with different knowledge to create a more comprehensive and complete picture of our business problem.

2.2 Practical implementations of big data

In general, big data enables organizations and governments to acquire new insights into their environment. The possibilities are countless and will grow in the near future. It will help organizations to comprehend business from different angles and unveil issues or occurrences that could not be revealed before. Probably one of the most interesting characteristics of big data is the ability to predict future scenarios with a very high accuracy. While this might be still in the distant future, actually having an impact on business, prototype systems already today are using predictive analytics to for instance predict areas in a municipality where crime are likely to happen at the very moment (IBM, 2014). The same technology also enables Google's self-driving car to anticipate the next move of other drivers in the traffic and adjust its own moves accordingly. The core of predictive analytics is the assumption that everything follows certain schemes. With enough data available it is possible for the data scientist and ultimately their databases to recognize schemes out of millions of petabytes worth of data.

2.3 Big data technology and software

Software has obviously a very crucial role in big data. It is the tool data scientists use to analyze data and helps to make data visual for the management. Some of the most popular software tools and software providers in big data are:

Apache Hadoop:

Apache Hadoop is the single most characteristic software tool in the field of big data. Originally developed by Yahoo and due to its open source license extended by Google, it is the foundation of big data in most companies. It is the software framework for the

storage and processing of data. The software is extendable by a number of modules to adapt it to the needs of the user. The software framework is mostly written in Java and C. Prominent users of Hadoop are companies like Yahoo, eBay and Facebook just to name a few (Hadoop.apache.org, 2014)

SAP

The German company SAP, widely known for its ERP solutions, also offers a variety of solutions related to big data. The offer is extensive and split up into different categories from analytics to services and visualizations. One of the interesting applications that SAP offers is Lumira. This application focuses on the visualization of data. It enables the user to create advanced figures and graphs by linking Lumira to the preexistent database. What makes this interesting is the fact that with this application, SAP stresses the importance of visualization of big data. With Lumira, SAP tried to provide a software tool with a simple user interface to enable staff without extensive training in IT to create visualizations of data that can be used to justify and pre evaluate business decisions (SAP, 2014)

IBM Watson

Just as the other leading business software companies, IBM also offers a range of big data solutions to its customers. The approach is more or less the same. More interesting is however IBM's Watson project, which is based on the developments of big data. As IBM states "Watson is a cognitive technology that processes information more like a human than a computer – by understanding natural language, generating hypotheses based on evidence and learning as it goes." (IBM, 2014). While the processes behind Watson are not solely based on big data, the core of Watson relies on large inputs of data. In fact, Watson accumulates data and draws conclusions out of these immense amounts of data. IBM claims that Watson processes data more like a human, which is basically the result of Watson being able to intake and analyze more data than ordinary computers. This demonstrates very well, like Google's self-driving car, the impact and advantage big data has. It literally enables machines to operate by themselves by being able to analyze increasing amounts of diverse data and interconnect the results.

Data visualization at Procter & Gamble

Procter & Gamble has realized the significance of good and reliable visualization in the business decisions making process and has introduced a revolutionary approach to data visualization. Business Sphere, which is owned by P&G, is a patented visual data environment. According to P&G, the system processes as much as 200 terabytes of data, presenting the data in all Business Sphere locations in the same way. Business Sphere locations, of which about 40 exist at the moment, are P&G offices around the world which are equipped with a special room (see picture below) that is specifically set up with visualization tools. The visualizations are around the world the same, helping managers to get used to one way of visualization and being able to easier collaborate across borders (P&G, 2013).

Figure 4. P&G Business Sphere conference room (P&G. 2013)

While most full analytics solutions by SAP, IBM, Oracle and alike require essential investments into IT, at this stage the basic big data infrastructure is open source and mostly available for free. However, this is not true for analytics or visualization software. Software providers have realized the increasing significance of big data and are currently trying to educate customers with publications of whitepapers and extensive information on the web regarding big data solutions.

The key to a successful implementation of big data in the organization is not solely based on having the technical capabilities to collect and store data. The visualization of data plays the most important role in management (HBR, 2013). Especially at this stage, data scientist are a scarcity and most business decisions are made based on experience and intuition. Software with a simple layout and the power to enable managers to use data for their decisions is here the success factor. At this point, most professionals with a big data background are career changers from scientific disciplines like social science, who

are familiar with the statistical tools to analyze large sets of data, however, they lack the management experience and mind-set to make strategic decisions.

The challenge for educational institutions and companies alike is to offer attractive programs and role for these individuals and find suitable ways of integrating them into the organizations (Economist, 2013).

2.4 Analytics methods

Big data and its analysis can consist of multiple very different methods of analysis. Some of the methods have been utilized already for decades in business, however, through big data technology they have become more accurate and analysis is able to consider a higher variety of parameters. The different analysis methods basically focus on different information that can be found within the data. The same data may be used as basis for all different analysis methods, however the results of the analysis may show a totally different picture, even though the same data was used.

Descriptive Analytics:
Descriptive analytics is the use of historic data to describe mostly past events and describe a given situation. In a business case, this may be the analysis of past marketing events, campaigns or demographics of a population. The major purpose is to understand and built a clearer overview of past events

Diagnostic Analytics:
Diagnostic analytics help to discover correlations, mostly by the use of visual tools. These correlations may trigger further investigation. As mentioned before, the results are mostly presented visually and do not aim to give a too accurate picture of a given situation. The major purpose is to unveil hidden correlations that may affect e.g. a market. This may be interesting in marketing, where diagnostic analytics can help to identify micro markets or certain patterns in customer behaviour.

Predictive Analytics/ Data Analytics:
Predictive analytics enjoys most of the attention in the field of big data analytics. The curiosity possibly derives from the fact that predictive analytics has the possibility to predict a future scenario with a very high accuracy. Predictive analytics are often used to predict sales figures and forecasts of different kind (Payandeh, 2013). However, they

may also be used in robotics. Theoretically, an industrial robot or mechanism with access to predictive analytics may be very useful in the prevention of accidents, e.g. in the automotive industry. The possibilities here are very vast and leave a lot of space for imagination because of the fact that this is one of the latest developments in big data technology

Prescriptive Analytics:
Prescriptive analysis is automated analysis of data that creates predictions and suggest decision options (Evans, 2012). This sort of analytics can mostly be found in industrial operations or inventory management. Prescriptive analytics follows clear schemes and offers a few decision options. This sort of analysis has been used for a long time already; however, with increasing amounts of data the offered options by the analysis are more comprehensive and able to consider a greater amount of parameters, thus making the information more accurate.

3 Big Data Challenges

Big data creates challenges not only in its implementation, but it also challenges lawmakers to find new appropriate limitations to this new use of data. Organizations will struggle in the first years of implementation to define roles for big data administrators and to encourage the use of big data in organizations. There are also challenges from a technological aspect, which we will discuss in this chapter.

3.1 Data quality

One of the greatest challenges in data management and utilization is the data quality. In order to gain valuable information from data, it is essential for the data to be of sufficient quality. In the age of big data, with ever growing amounts of data, organizations face challenges in maintaining a sufficient level of data quality. In fact, an entire industry that deals with data management and quality assurance has emerged from this important need to assure the highest level of data quality (Forbes, 2014). In this section we will take a closer look at the theoretical requirements to data and steps that organizations have taken in order to maintain a high degree of data quality.

A first prerequisite for accurate data is of course that the data collection is set up correctly. This requires the technology used to be configured according to the data collection

and information needs of the organization. A good data quality is defined mainly by the data's suitability for its purpose. Only sufficient data quality can allow analytical departments of an organization, like Business Intelligence or Business Analytics, to conduct reliable analysis and reports.

Speaking about big data quality, the terms volume, variety and velocity of data are crucial elements of reliable and appropriate data. Apart from these, availability of data, cleanliness and correct data contribute significantly to the value and usability of data. Especially the cleanliness of data represents a major challenge (Int. Assoc. for Information and Data Quality, 2013). Often the data mapping, combining different sets of data and structuring data, leaves space for error. In smaller systems, manual labor, mostly done by a Master Data Management team, as a corrective action can help to assure a good level of quality. Big data, however, is in volume simply too big to be under manual surveillance. Systematic solutions, provided by major players in the data industry like Teradata or IBM have developed solutions to tackle this problem (IBM, 2014).

In general, data quality is more dependent on the business culture, rather than the IT systems. Poor data quality is often the result of poor data entry. Therefore it is important to promote a detail oriented approach within the organization to assure a high and consistent level of data quality. IT solutions are able, to a certain extent, to systematically improve data errors. In many organizations this can already be achieved by limiting the scope of manual data input. In practical terms, free text fields in data sheets are at a significant higher risk for errors, such as misspelling or typos, than fields with predetermined values.

The implementation of a Master Data Management team can help to minimize fluctuations in data quality as well as to create a central reference and maintenance point for data issues (Microsoft, 2006). As mentioned before, the industry has already recognized the challenges of data quality. Standards like the ISO 8000 have been developed to create guidelines and assure a vigorous minimum of quality.

For years, quality management has been an important part of operations. In the last few years, quality management has also reached the data management of the business. The exponentially growing amount of data has made it challenging for organizations to maintain their current approach to data and quality. The realization that new approaches are required in the future is making it slowly into organizations and has reached a stage of

early adopters. It is a process of transition. The easier part is the implementation of the required IT. Significantly more difficult is the work that change management has to do in terms of changing the company culture and educate a literacy in data treatment.

3.2 Data privacy and security

In recent years, the public has come more aware of the capabilities data has and how it is affecting people on an individual level. The NSA spying affair in 2013 is one of the contemporary issues related to large amounts of data that have been stored with often questionable intentions. Governments, as well as private institutions possess large amounts of data, which equals to private information, which, if analyzed extensively, may reveal information about individuals that is considered to be private.

With sophisticated tools, data scientist are able to filter data to an extent that can narrow down information about a single individual within the group of millions of people that data was collected on (HBR, 2013). Over time, this problem will grow to be more severe, as mobile devices, as well as stationary equipment in households are increasingly sources of large amounts of data and connected to a network with the capability of storing data.

With the amount of data that is collected rising and storage capacity growing in the future, it may be possible to anticipate every individual's behavior with high accuracy. If the current growth rate of data maintains as high, this may be, according to a Stanford Law Review paper, a reality soon. Doubts about the security of big data exist.

> Big data is really the amalgamation of little data inputs — information about people, places and information collected by sensors, cell phones, click patterns and other data-generating mechanisms. These data inputs are collected by commercial and government systems, for big data purposes (Ferguson, 2013)

Most data is collected invisibly, without the knowledge of the individual. Access to the data however, can represent a major threat to the privacy of a group of people or an individual. Individuals should still have the right to be the owner of their own data. This is especially the case for data like geo track data, phone records, buying records and especially health data, which allow deep insights into the personal life and habits. The possibility of misuse of this data poses a major threat to the individual.

From the viewpoint of an organization, big data does not only allow the organization to assess its business environment and customers, employees within the same organization might face new ways of data surveillance. Through big data, employers have potentially the capability to monitor employees in detail. Performance measurement could lead to more pressure and a hire and fire culture, threatening the stability of organizations. Big data is already used widely to study the behavior of employees in order to increase productivity. Famously, a call center used big data to increase efficiency of staff by 10%, as they discovered that call agents are more productive when chatting with each other during breaks. The result; the employer scheduled group breaks instead of solo breaks, which eventually leads to the employees mingling with each other, rather than taking breaks on their own in solitude. (Silverman, 2013). This could lead to a more sophisticated twist to Taylor's Scientific Management theory[3]. Big data enables new possibilities to analyze work and can, just as Taylor's Scientific Management theory, put workers under unproportionally high pressure to perform tasks faster.

This development demands for sophisticated security measures to protect individuals from the possibility of misuse of their data in private life, as well as in the workplace. Furthermore, there will be a need for regulations concerning the use of data. Governments and private organizations have to follow rules that assure the integrity of individuals. The authors of the Stanford Law Review article (2013) suggest the development of a concept of big data ethics. Curbs on the use of personal data, combined with new technological options, can give individuals control of their own information, according to the report, while permitting important data assets to flow relatively freely (Lohr, 2013). It is a fact that big data will allow a better insight into the actions of individuals, as well as the possibility to predict actions before they actually happen. Without regulations and laws that control the use and the ethics of big data use, the harm to society might be more severe than the gains in productivity

3.3 Big data talent

The changing business environment due to the emerging of big data technology has an increasing effect on companies' talent acquisition demands. New technologies and approaches require new skillsets and levels of software literacy. Business Schools around the world have not yet been able to provide graduates with an extensive knowledge of

[3] Management theory that analyzed and synthesized workflows. Determining "the one best way" to perform a certain task.

big data and analytics. Courses with a specialization in analytics are a scarcity. Only recently the top management schools in the United States have started to implement programs with a focus on business analytics.

> Yale's School of Management has a Centre for Customer Insights, with IBM among the funders. Wharton has a similar Customer Analytics Initiative. Kelley School of Business at Indiana University not only launched an MBA in business analytics last year, but has signed a deal to offer something similar at the Indian Institute of Management (Economist, 2013).

According to a McKinsey report, in 2018, alone the United States will face a shortage of 140,000-190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze big data. The big data evolution demands for a new breed of professionals; data scientists (Davenport, 2013). Data scientist play a new key role in organizations, with the skills and capabilities to make discoveries in large data sets, gaining information crucial for businesses. Currently the focus in big data technology is usually on the software tools that give access to the data and make the structuring of data possible, allowing its analysis.

4 Industry Case Examples

The following case examples demonstrate how organizations have been implementing big data into their operations in order to increase efficiency or focus on new market segments. The presented cases are considered ground-breaking and do not represent the general stage of development in most organizations.

4.1 Big data in agriculture

Unlike the common believe, the agricultural industry is a high tech industry, utilizing data in its operations for a number of years already. High tech fertilizers are used, as well as weather forecast data to align harvest plans with the external conditions. The agricultural industry is highly dependent on external factors that to a high extent cannot be manipulated by human force. Therefore, deliberate planning is of high importance. The importance of the agricultural industry for the human society is equally important. The UN estimates that the world population will grow to 9 billion by 2050, especially in the Asian economies the demand for food is rising in proportion with the increasing standard of

living. The UN organization for agriculture estimates that the global food production has to increase by approximately 70% in the next 35 years to be able to cope with the increasing demand. At the same time, farmland is becoming a scarcity, as it has to dominate with land used for the more lucrative bio fuel production (United Nations, 2013).

Big data technology can help to increase productivity in agriculture and minimize waste. More sophisticated analytics, based on larger data inputs enable organizations to provide more accurate and detailed weather forecasts, resulting in an improved planning ability for farmers. The better data insights can for instance help to determine optimum harvest dates, growth rates and time windows for ideal fertilization. In general, the more data and analysis available, the more precisely a farm operations can be planned. In the dairy and beef production, big data can be utilized to give better insights about the cattle's health conditions. A Texan start-up developed a solution that enables farm managers to get detailed real time information on the cattle's health, as 40% of dairy cows get ill yearly, causing, according to the US department of agriculture, a loss of 5 billion US dollar a year and globally up to 60 billion US dollar a year. The data that is gained by cows swallowing a sensor, originally developed for military purposes, which whenever a cow walks past a wireless receiver transmits data about the cows vitals, acidity in its rumen, hormone levels and temperature. The collected data is uploaded to a cloud based IT platform giving farm managers detailed information about each individual farm animal. SMS and e-mail alerts are sent out whenever an indicator hits a critical level. This can help the managers to intervene before health issues for cows arise and in general helps to act more promptly. The data furthermore enables to set up industry benchmarks and increase overall productivity (BBC, 2014).

Another factor that has a mayor influence on the global food supply is the fact that, due to the increasing complex supply chains and long transit times, 10%-15% of transported food products get spoiled. The causes are mainly temperature fluctuations and humidity. The result is a global loss of approximately 25 billion US dollar worth of food. (BBC, 2014). This is not only a problem in the food sector, but strongly affects the life science sector as well. DHL, one of the world's leading forwarders, has developed a solution to cope with this issue, which will be discussed in the logistics case study of this research.

4.2 Big data in logistics[4]

As mentioned earlier in this research, transit is a crucial point for various sensitive goods and causes yearly billions in losses due to inappropriate handling or environmental issues. From personal experience, in most cases temperature and humidity are factors that can cause damage to transported goods. Especial the life science industry has a range of products that are sensitive to temperature, humidity and even vibration caused during transport.

One of the leading forwarding companies in the industry, DHL, has identified this issue and is one of the first organizations that are able to provide a solution for sensitive goods. DHL Thermonet is a data based solution by DHL that enables permanent monitoring of important indicators during transport. With Thermonet, DHL is able to provide reliable cold chains and can assure adequate treatment of goods. Container and packaging are equipped with sensors that monitor in real time temperature and humidity in and outside of the packaging. The data is uploaded to the Life track IT platform of DHL, which is accessible by customers to monitor not only the temperature and humidity levels, but are able to track via GPS the location of a single container worldwide. The application is able to send alerts whenever parameters exceed the set limitations and can help to assure a certain quality standard during transit. The data on the other hand helps DHL to identify reoccurring situations where certain temperatures are reached or delays in transit are most likely to occur (DHL, 2013)

Experience shows that this is, however, only one side of the coin, since the data that is stored primarily to offer the customer assurance of proper handling of the goods, helps DHL on a second level to get better insights into their methods of transport and with the help of predictive analytics to develop more efficient transit ways. A sophisticated system that collects data from all moving parts in DHL's operations can for instance be used to react in real time on delays of certain legs of a transit. Changes can be made before delays occur, customers can be notified timely and errors that reoccur can easily be identified.

Contracts between a freight forwarder and large clients often contain clauses about a minimum of timely deliveries. The leading freight forwarders are usually able to promise

[4] This part of the research is to a large extent based on the researchers own professional experience

around 95% timely deliveries to its customers[5]. Whenever delays occur that are not due to the performance of the freight forwarder, but have its origins in external factors, the 95% limit is not considered. Big data and constant monitoring of shipment movements enables the freight forwarder to track exactly where during transit errors occur and ultimately the forwarder can notify the customer whether the delay occurred due to problems on the freight forwarders sight of responsibility or whether it is actually the customer that didn't meet a deadline. Often shipments are stopped at customs, due to missing or incomplete paperwork, which more than often is the responsibility of the client. In that way big data can also help the client to identify parts of its own operations that may be subject to a higher error rate than other departments.

In general, big data supports logistics operations by showing real time movements of shipments by identifying error hot spots. Predictive analytics can help to prevent delays, determine bottle necks and provide more accurate transit times. The commercial sector is naturally the first to profit from big data technology. However, the current rate of development will certainly affect logistics operations for the private consumer as well. A possible example is the high rate of lost luggage on airports. According to a SITA baggage report, in 2012 over 26 million bags have been mishandled on airports. As sensors, like the popular RFID[6] sensor, continue to become more affordable. Airports will be able to track passenger's bags in real time in the future. Alerts can be sent out whenever a bag is directed into the wrong direction. This will help to lower the amount of mishandled baggage to a minimum, increasing overall customer satisfaction of airlines and airports around the world (SITA, 2013).

[5] Own estimate based on professional experience
[6] Non-contact radio frequency data transmission (see appendix)

4.3 Big data in retail

The retail industry is an industry that is very like to gain great advantages through big data. Big data will not only help retailers to identify new markets and business models, but it will also have a significant effect on its operations, as demonstrated by the previous logistics case.

4.3.1 Amazon's anticipatory shipping

One of the most innovative companies, not only in retail, but increasingly in technology is Jeff Bezos's Amazon. The largest retailer in the world is slowly moving from a business model solely based on e-retail to a data company. Well known for its operative efficiency, in late 2013 Amazon made the news with a new patent that it registered (United States Patent and Trade Office, 2013). The Wall Street Journal, which was first to discover this new patent called it "Amazon wants to ship your package before you buy it" (2014). This title describes this newly registered model quite accurately. At the core of this method sits predictive analytics.

As Amazon is online based, it has extensive data insights into its customers and their activity on the Amazon webpage. Big data technology now more than ever enables companies like Amazon to actually analyze the data. The result is that Amazon plans to implement a system, which would anticipate items that are likely to be bought by a customer in an anticipated location, in order to start the shipping process of the item before the customer actually purchases the item online, thus being able to deliver items faster.

Figure 5 demonstrates the process of this method. The system anticipates that a specific item is likely to be bought in a specific location. The shipment process starts and the item is shipped towards the anticipated location. During transit, as the customer makes the actual purchase online, the preassigned address is specified to the name of the customer that bought the item online.

Amazon describes the method in the patent document as follows:

> A method and system for anticipatory package shipping are disclosed. According to one embodiment, a method may include packaging one or more items as a package for eventual shipment to a delivery address, selecting a destination geographical area to which to ship the package, shipping the package to the destination geographical area without completely specifying the delivery address at time of shipment, and while the package is in transit, completely specifying the delivery address for the package (Spiegel et al, 2013).

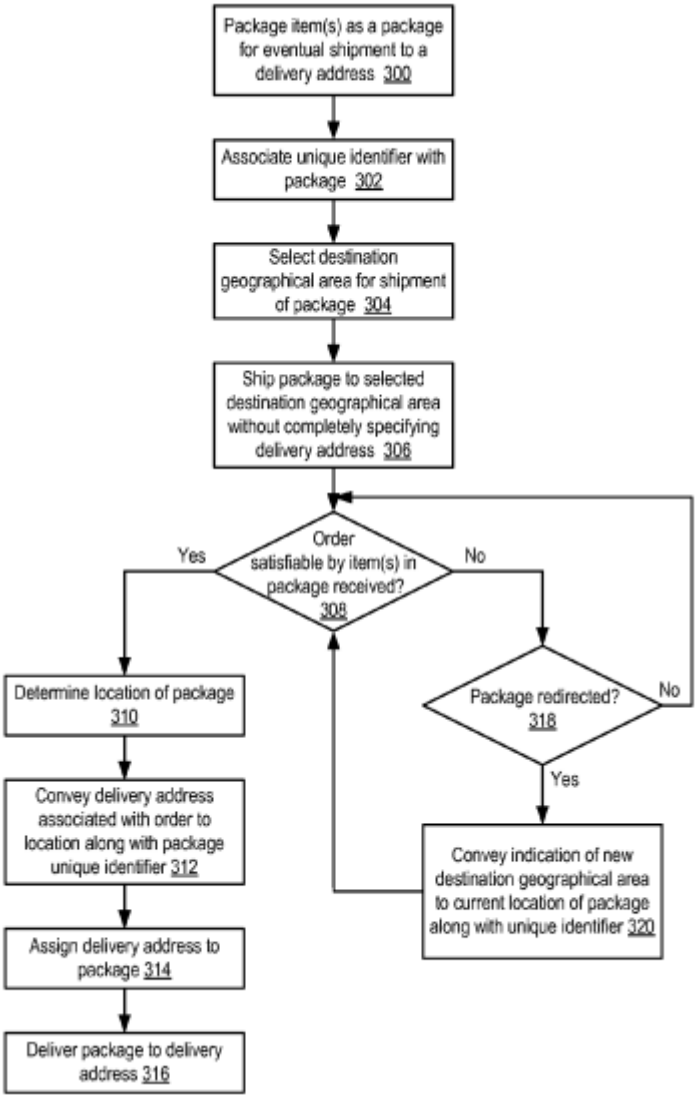

Figure 5. Amazon's Anticipatory Shipping (US Patent Office, 2013)

According to the Wall Street Journal (2014), Amazon has not yet commented on whether it is already applying this method or is planning to do so in the future. However, this method seems to be advantageous if we consider the example of a bestselling book as

an item in question. If we consider e.g. an apartment block, with hundreds of potential Amazon customers, the likelihood that at least one of the residents is going to purchase a recent best-selling book is quite high. Therefore this method would be able to minimize the waiting time for customers to a minimum, without physically moving the book any faster. The advantages of this method solely derive from information that Amazon has about its customer and not by making operations faster really (WSJ, 2014).

4.3.2 Recommended items

One of the most famous application of big data in e-retail is the suggested items panel on many e-retailer`s website, which suggests items to the customer depending on previous data. This system is already since 2003 in use and based on collaborative filtering. Amazon used and developed this system, to perfection and many other retailers followed. A similar system, based on almost the same algorithms is the “people you may know” feature on LinkedIn, which frequently suggests contacts you might already know, based on the information available on a user’s profile. This feature, which is now responsible for more than half of the connections in LinkedIn, was introduced by what is known to be one of the first data scientists in business (HBR, 2013).

4.3.3 Customer loyalty programs

Customer loyalty programs, such as bonus cards, have established themselves for years in the retail business. From a data perspective the benefits of distributing these cards are obvious. Customers give insights into their buying behavior, which then can be used by the retailer to customize their operations. For a long time however, companies have not been able to fully benefit from all the data inputs, as the technology was not able to cope with this large amount of data.

An organization that has been doing particularly well in utilizing this program is Sears in the United States. The department store is still mostly brick and mortar based, however has developed a sophisticated customer loyalty program that gives the organization great insights into their customers. The main benefit of the program, compared to regular programs, is that Sears’s best customers actually know that they are the best customers, making it more of a reward program. The system is in general very focused on the individual shopper. Competitors like Walmart, who do not have a loyalty program, gain their

customer information through their Walmart credit card service. However, this is not as detailed as the award winning Sears program (Forbes, 2012).

4.3.4 Big data touch points in retail

Retail is one of the industry sectors that is most likely to benefit greatly from a big data implementation. As shown in the figure 6, the number of possible big data touch points is immense.

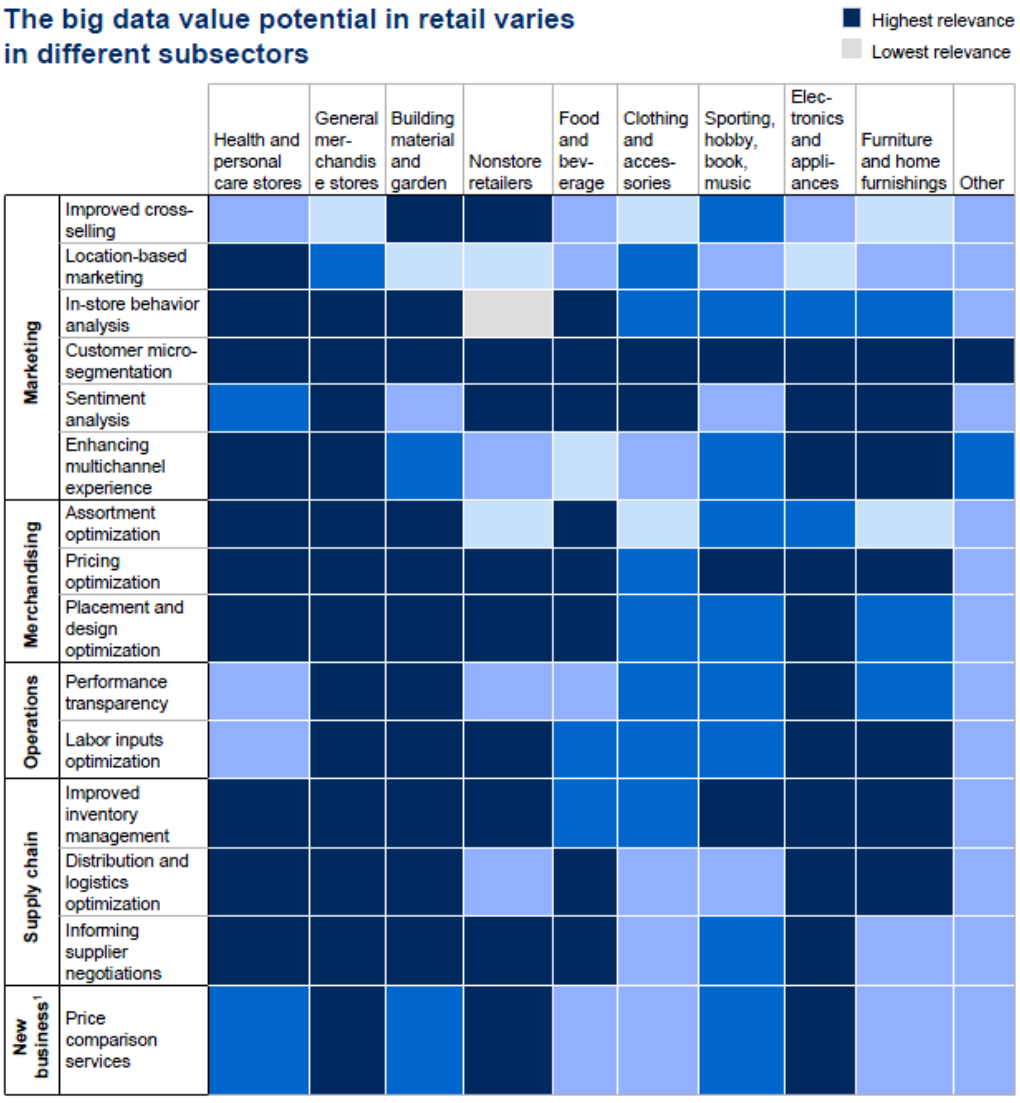

Figure 6. Big data relevance in retail in different subsectors (McKinsey Global Institute, 2013)

The reason for the vast amount of touch points in retail lies in the nature of the business. Multiple supply chains as well as different business functions can all benefit from big data

individually. The possibilities in retail, however, can also be a good example and guideline for the future implementation of big data in different sectors.

4.4 Big data in healthcare

The big data evolution in the healthcare sector is one of the most anticipated impacts of big data, as it has the capability to impact the life of the general public significantly. Improved healthcare services can reduce costs and contribute, worldwide, to improved standards of living. There are many scenarios of a possible implementation of big data into healthcare, the most practical way of using big data is the analysis of historic patient data. Figure 7, demonstrates a few of the possibilities in healthcare created by big data usage in the US healthcare apparatus.

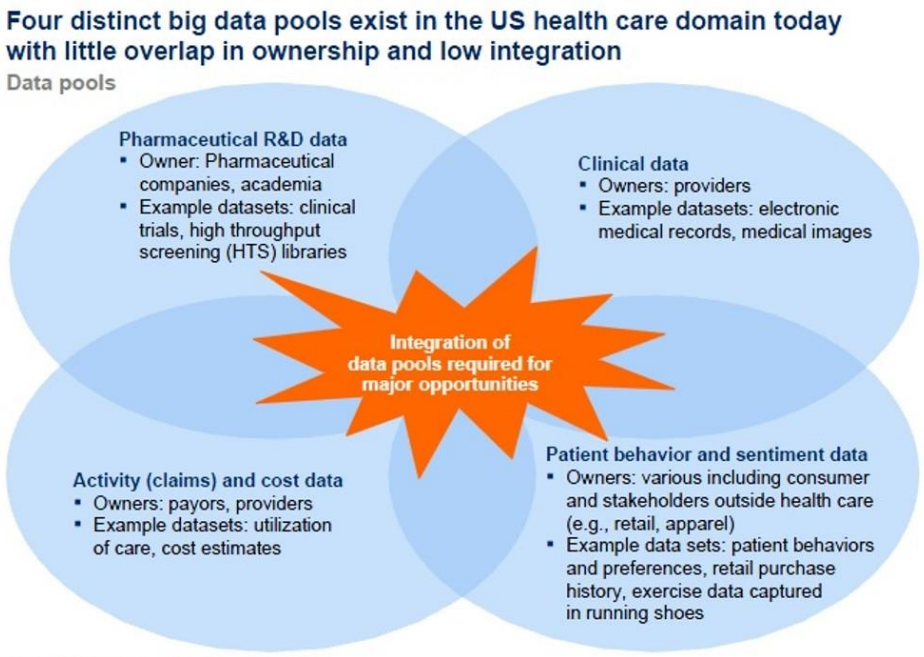

Figure 7. Big data in the US healthcare sector (McKinsey Global Institute Analysis, 2013)

Having the accurate information on a patient is an important factor in healthcare. Usually a general practitioner has a better knowledge of a patient, compared to doctors in an emergency department of a hospital. However, both usually face a problem of having only access to scattered information once a health problem occurs. The diagnosis of a certain condition and the necessary treatment is often difficult and multifaceted. Big data can help to narrow down possible causes for health problems and fill information gaps to provide a more accurate diagnosis.

One way of implementing big data into healthcare is rather simple. It is the creation of a detailed patient database, which includes historical data, current patient data and information on the circumstances and behaviors of a patient. Big data analytics can enable a practitioner to have an overview of a patient's history, including his frequency of hospital visits, long term symptoms and conditions as well as habits like regular exercising etc.

A large number of diseases can be treated better, if diagnosed early. Data analytics can help to identify early symptoms of common diseases, not only by looking at the individuals health indicators, but also by comparing patients with similar conditions and life circumstances. A cloud based system is able to compare millions of patients in seconds by identifying similarities and abnormalities alike. Furthermore, the treatment can be highly customized and compared to successful treatment of patients with the same symptoms in a very detailed manner.

Risk patients may be equipped additionally with e.g. wristband sensors that frequently update patient`s vitals in the database. A system like that can be combined with health alerts, which are warning patients and practitioners once a health parameter reaches a dangerous level. Emergency doctors would benefit from having a complete overview of a patient by having access in seconds to crucial patient data. Obviously, the main benefits of big data in healthcare are improved information flow and time savings, which are able to save lives in this industry (IBM, 2013).

The healthcare industry and big data will create a range of possibilities for businesses to benefit from these developments. New business models around patient data will be developed and increased information flow will improve a number of healthcare services. A challenge in this industry, more than in other industries affected by big data, will be the aspect of privacy. Patient data is very sensitive and each patient will have to give consent to the use of his data in databases. Therefore, this industry will require a sound regulatory framework to prevent misuse of data and to prevent creating a transparent patient. Big data will enable practitioners to give more evidence based treatment, rather than treatment according to intuition. Big data in healthcare is a tool, which is supporting the decision making process of healthcare specialist by giving better insights into patients, clinical trials and pharmaceuticals.

According to a McKinsey study in 2013, it is suggested that the healthcare sector has to redefine a value framework in order to maximize the profits in big data technology. Patients must be encouraged to right living and to play an active role in their own health by having access to information that helps them to maintain a healthy lifestyle. At the same time, all caregivers that are involved in the patient's health, have to have access to the same information at all times to consequently maintain a high level of services. Data can also help to determine the right provider of care services. In order to save costs and time patients need to be treated by specialist with the right skillset. Overqualified or underqualified caregivers are a cost factor and at the same time may prevent timely treatment of patients that are actually in need of a higher qualified practitioner. The benefits of big data do not only have an effect on the physical patient care, but research and development can benefit greatly from better information flows and more patient focused products (McKinsey, 2013).

4.5 Big data across industries

Big data technology is relevant to all industries, especially but not exclusively data and software companies. However, traditional sectors are increasingly benefitting from big data technology as well. In general, big data is still in an early stage and companies have just started to implement big data into the business. This increases the impact of big data for those, who decided to utilize the technology already in its early stage and have gained a competitive advantage towards other organizations in the industry.

Apart from the example cases in previous chapters, other industries that are forward thinkers in the field of big data are the finance and insurance sector, which uses sophisticated analytics to determine return on investments and to gain detailed pictures of future scenarios. An area affecting all business across industries, is big data marketing. Big data enables companies to get a unique perspective on their market and the possibility to target micro markets, which have been invisible until the age of big data. Micromarketing plays an important role especially in online business (HBR, 2013)

In general, big data technology facilitates the decision making process in all industries by giving new insights into business issues. Analytics provide the information that supports decisions and help to gain a better understanding of the business. On the other hand, big data and the application of the technology creates entirely new industries centered on big data business. New business models and services emerge around big data

to facilitate its use. Big data influences products ranging from intelligent kitchen equipment to highly complex analytics tools, which shows how significant the role of big data is to the entire economy. Industries are based solely on data. The real change, however, will be visible once the technology has actually established its role in the society.

5 Big Data in Management

The implementation of new technology is usually quite difficult in most organizations. Some organizations are known to be able to cope better with change than others (Osty, Uhalde, 1999: 299-313). This is highly dependent on the company's culture as well as the change management. The implementation of big data into the everyday processes of organizations could be as challenging as the implementation of ERP systems or the change to a paperless office.

Errors are common in the configuration of the systems, however the most challenging part is usually the change in business culture. The corporate culture of an organizations is of fundamental importance in introducing new technology. The corporate culture can be quite complex and consist of many elements, like the model of the cultural web demonstrates (Johnson, Whittington and Scholes, 2012). Unsuccessful implementation of technology into the business culture can be the cause of more harm than benefits created by the new technology. In this part of the research we will mainly focus on the role of big data in the decision making process. To understand the role of big data in the decision making process we have to understand what kind of decisions are made in business and in more general level, how decisions are made.

5.1 The simplified decision model

In Management we usually face two kinds of decisions. Operational decisions and more complex decisions like in strategy development. Operational decisions like in logistics can be automated to a large extent with the help of big data. The calculation of reorder points in inventory management follows usually a structure and is therefore easily understandable by a computer making decisions based on data. A rather simple algorithm with access to the appropriate data could optimize the economic order quantity and contribute significantly to costs savings in inventory management. On the other hand, strategic decisions are often more complex and require a deeper understanding of the situation

and the ultimate goal that decision is supposed to achieve. Therefore we need to understand the theory of the decision making process. There are a few psychological theories about how decisions are made, but in this context we can simplify the decision making process into 7 steps that are taken in order to make a decision (Brown, 2007)

1. Identify a situation which requires a decision
2. Gather information to base your decision on
3. Identify the different options
4. Estimate outcome of all options
5. Compare different options/scenarios
6. Make decision
7. Evaluate result

These steps are usually taken when making a decision. Big data can have an effect and influence all of these steps. However, at this point most decisions are based on experience of the HiPPO (Highest Paid Person's Opinion). Experience and intuition are most likely the deciding factors in top management decisions. Data is often used only to evaluate the result or to identify a situation that requires a decision (McAfee, 2013).

An approach to the decision making process by supporting the process with insights from data can help businesses to make better decisions. In fact, a research by Erik Brynjolfsson (2011) has shown that companies that have implemented big data as decision support are outperforming competitors without big data technology in business performance, measured by asset utilization, return on equity and market value,. However, this is only one of the very first pieces of research that have studied the relation of big data and more successful decisions. Nevertheless, first indicators show a positive tendency. The question remains how exactly big data might be used in the decision process to make better decisions.

If we look back at the simplified decision making process model and its seven steps we can identify a number of opportunities that would allow to use data technologies as decisions support. Firstly, big data can help to identify situations that require decisions more easily. Analytics systems have the capability to analyze large sets of historical data and compare it to current outputs and to find situations of similar nature that have required a business decisions before. Furthermore, the mass of data and especially its variety and

multiple origins within the business can improve a manager's overview and understanding of the situation. As mentioned earlier in this research, a system like the one used by P&G, visualizing data, can aid to develop more sound decisions and strategy.

The key strength of big data compared to the use of data in the past decades is the possibility of highly accurate predictive analytics. This enables decision makers to make experiments with their decision without actually having to carry out the decision. Software and data are able to estimate scenarios and therefore are a good place to conduct experiments in a safe environment. This is especially crucial in comparing different decision options.

Summarizing big data as decision support, data helps mainly to get a better understanding of the current situation and allows to calculate scenarios. Compared to decisions based purely on experience and intuition, big data based decisions are relying on facts rather than feelings and estimates by a manager. One might argue however, that competitors and the business environment are of an unpredictable nature and data might not consider this element of unpredictability. Research suggests though that human behavior and the behavior of the market follow a quite predictable pattern over time (Boyd, Crawford, 2012). The key to identify patterns is to look at the very long time spans, which is now possible with big data. Companies are able not only to analyze their performance in the last few years of business, but in theory they have the possibility to analyze their development over the entire time of their existence. Once old data, which may have been saved in hard copy, is transferred into a digital format, the data could be utilized by analytics software

5.2 Big data in a decision making process model

In the following chapter we will take a closer look at the findings of the primary research that was conducted in face-to-face interviews for the purpose of this research. The interviews were of unstructured nature and used the help of guide questions to initiate a discussion related to big data. The primary research ultimately consists of two interviews that have been conducted with professionals working in the field of big data, with extensive experience of each more than 15 years in the industry.

The first professional, from this point on referred to as Interviewee A, is employed by one of the leading German telecommunications companies with a portfolio that well exceeds

Germany. The company has a global presence with a strong market position in all important markets like the United States, Europe and Asia. The product portfolio of the organization includes IT infrastructure services, IT consulting, mobile broadband and landline services. Formerly a state owned enterprise, the organization is now privatized and benefits from its previous role as state owned enterprise. Regardless of the fact that the company is active in one of the most innovative industries, the structure is rather traditional and very hierarchical and typical for ex-state owned organizations.

The second interviewee, referred to as Interviewee B, is as well active in the It industry and has had a consulting role in a number of small and medium sized firms and advised organizations on their strategy and product development. Both interviewees have an academic background in mathematics. What follows is a reproduction of the statements made by the experts during the interviews.

According to Interviewee A, the tendency towards implementing big data in his current organizations is twofold. The company has a history of utilizing data from various sources to gain insights into the business environment. Interesting is the fact, that the company is almost exclusively the owner of the broadband and landline infrastructure in Germany, which is a result of the previous state owned role. This gives the company the advantage of having access to detailed information about the competitor's use of the infrastructure in Germany. Therefore, data analysis has always played an important role in the organization. Big data, as successor of the traditional data applications used, has a very strong base. Especially in the mid management of the organization, decision makers are likely to rely on big data and implement it into their practices.

The company culture however challenges this positive attitude towards big data. The organizational structure is rather complex and still follows a certain hierarchy and information/power channels. Unlike many other telecommunication companies, which follow very flat hierarchies. Change has from experience always been difficult within the organization. The same is true for the implementation of big data as decision support. The consideration of historical data is well established, new technologies have, however, not yet made justice to their capabilities. The culture, as well as the attitude, of often more senior managers, still drifts towards making decisions based on experience rather than modern analytics.
The interviewee however confirmed that development is visible and that the big data revolution has also already started to spread into the top management ranks.

The second interviewee has experience from working with a number of companies that are different in nature. Working with a larger number of different companies his experience still underlines the statements of Interviewee A in general. The company culture has a significant influence over the acceptance of big data technology within the company. Long established, bigger organizations can afford to take a back row set and watch big data technology develop in front of their eyes. Smaller organizations are required to be more innovative and take chances with higher stakes in order to survive in the market. Therefore, smaller organizations and especially start up as very willing to implement latest big data trends into their decision making. In fact, most startups consider themselves as leading in the field of big data. Big data is on eye level with other previous technologies like the internet and possibly utilized to the maximum in order to gain competitive advantages through technology.

According to both professionals, big data technology will create a lot of opportunities for businesses across sectors. Nevertheless, it will be difficult to say what the effects will be at last. There is an element of unpredictability and neither is willing to make detailed prediction about the future. Crucial is however to keep an eye on these big data technology trends, as well as other new technologies and not fall behind in implementing it. To what extent big data technology should be implemented into the business has to be decided by looking at the individual case. Capital and labor skill are not unlimited, therefore all these factors have to be considered carefully. The capabilities of large corporations are unmatched by the very limited assets of fresh start-ups.

The greatest challenge of big data will not be privacy or any other regulation. The challenge will be to extract the main benefit of big data and to identify the key strength earlier than the competition. This is true especially for small businesses. Large organizations can afford to watch developments happen and acquire knowledge and development if necessary. The willingness to change is another crucial factor in implementing big data. Large corporations rely on a lot of decision makers that all have to be convinced. Small organizations are more flexible and are able to react to a changing IT environment more quickly.

Considering the findings gained by the primary research and the knowledge gained by analysis of secondary literature, the model in figure 8 could demonstrate one role that data plays in the decision making process of today’s organizations. We see the traditional

approach of combining experience and intuition to come up with a business decision. In this case, the decision maker draws a conclusion from similar cases and experiences and bases his decision on the fact that his previous experience is an equivalent to the current business situation. The contesting approach is a decision based on data analytics. Data is processed into information, which could be of predictive nature or just an analysis of historic data. In this case the decision maker would make his decision based on the suggestions provided by the data. These two approaches are both influenced by the corporate culture of an organization.

The culture is the deciding factor whether more decisions are based on experience or data. Innovative organizations that have equipped their human capital with data scientist and data analysts are naturally more likely to choose more data based decisions. Especially operational questions can best be answered by following structured rules as unveiled by the data. However, at this point it is suggested that the corporate culture should equally consider both ways of making a decision. Promoting a culture that combines the experience with extensive data analytics is, especially when dealing with complex issues, the way with the greatest benefits. While data allows the automation of many decisions, more complex issues should be approached with experience, intuition and data.

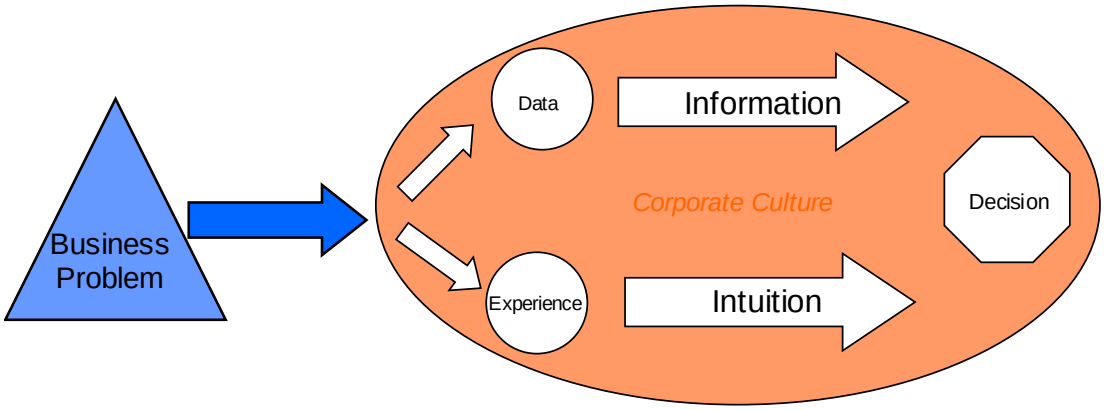

Figure 8. Data based decision model

The quality of data as well as the skillset of the data scientist may have a great impact on the representative value of the information gained by data analytics. A scenario in which a team of data scientists and analysts utilize the data to make decisions and compare the data to the experience and intuition of experienced managers would be ideal. However, in most organizations data analytics has not yet reached such a level of recog-

nition. As Eric Brynjolfsson points out, a combination of computer analytics and the human mind is able to beat any decision that is based solely on a computers calculation or a human's experience (2012a).

They key to success is to understand big data as a management tool, rather than a standalone solution. Therefore the key skill of the future is to efficiently make use of the technology, rather than to race or challenge the technology. Training and reimplementation of changed processes have to contribute to the adaptation of the change towards a more data driven decision model. Even with all the technology at hand, the culture has to give data the required credibility to be seriously considered in the decision making.

6 Conclusion

Finally we can conclude the findings of this research. To understand the advantages of the application of big data a very basic understanding of analytics is required. We see that the end user analytics tools play an important role together with the visualization of data in order to enable management and non-technical staff to be able to make decisions based on data. Within the organization there is also a need for educated data scientists and specialist that work on a daily basis with the preparation of data and that are able to understand the business needs of the organization while contributing to the success with their technical skills.

The challenges of big data lie within the organization, as well as in the environment. Regulatory laws and privacy concerns will have an effect on how data will be used and may limit the usage of data from a business perspective to protect the private individual from a breach of privacy. When misused, big data has the potential to threaten the privacy of an uncountable amount of individuals. As society has to redefine its understanding of a society with big data technology, a regulatory framework has to make sure that the use of big data stays within boundaries. A new ethical code has to be developed that defines the use of big data technology from an ethical point of view.

Within the organization big data has to justify its purpose as management tool and challenges the corporate culture to redefine management decision making. Depending on the corporate culture, big data will find its way sooner or later into management. Smaller businesses are generally more flexible and adapt easier to new technologies and new ways of working.

However, in terms of big data technology, the most innovative organizations can be found among large corporations. As stated before, organizations like General Electric and P & G are leaders in big data and invest heavily into this technology. The example of Amazon shows that they are in particular very well in developing new business models with the help of big data and increase the efficiency of their operations significantly. The advantages are obvious, better analytics allow companies to understand their business environment better and make more sound decisions that will eventually lead to a better position in the market. As pointed out by interviewee A in the primary research, information is the most valuable good an organization can possess to build competitive advantage.

These organizations lead the way for other small and medium sized organizations that will try to take advantage of the findings caused by the research investment of the industry leaders. As more and more well-known organizations rely on big data, other organizations build more trust on data technology and start to apply it to their own business. In addition, big data tools will become more sophisticated and user-friendly, which allows also start-ups to utilize this technology more efficiently.

The significance of big data technology and its implementation can be compared to the implementation of sophisticated ERP systems of the latest generation. These systems have improved operational efficiency significantly and have become indispensable in operations. Big data adds the advantage that it is a wider range within the organization. Analytics can be utilized in merely any department from HR, Operations to Product Development. The possibilities to support the business in a department with big data are numerous, including simple statistics as well as complicated predictive algorithms.
Big data goes however beyond this, as it creates the opportunity, like the internet, to reach customers differently and alter business models. This makes it a technology with a strong impact on all kinds of businesses. One could only imagine how an organization would operate today without utilizing the internet in one or the other way. The opinion leaders and interviewees see big data at least as significant as the internet. This alone justifies further research and development into the field of big data.

Big data is without question one of the most enabling technologies of the coming decades. The possibilities it offers already in its initial stage of development are vast and wide reaching. First indicators show that big data will be the cause for the development

of new business models, data based industries, new ways of measuring performance and parameters that could not be measured before.

The decision model in chapter 5.2 demonstrates how the corporate culture plays an important role, if not the most important role, in the implementation of big data as decision support. We have already discovered the advantages of big data and the challenges to the company culture. The model can help to identify the culture as factor when implementing data into the decision support of management. Data itself can only be useful when analyzed correctly and considered in the decision making process as reliable source of information. Data and the information gained should not challenge experience, but rather underline it. In an ideal situation, data helps to identify somewhat hidden facts and therefore helps managers to apply their experience and knowledge more appropriately. Furthermore, data can also help to understand one's own experience better. Drawing conclusions from previous data and relating it to the experience from past situations can help to build a better knowledge base.

In general, the society and businesses have to actively look to apply the technology to their benefit. In the past, companies that have started to work early with inventions like the steam engine, the computer and the internet have been able to maintain a very strong position in the industry. Big data is a technology that can be the source of very strong competitive advantages. Companies like Amazon, are already collecting patents in the field of big data technology and are sure to benefit from other companies that are interested in maintaining a competitive position in the market.

Acceptance of the fact that decisions are made differently are a basic requirement of a successful future with big data technology. However, companies and individuals cannot blindly follow the technology and trends. In a fast paced technological world not all developments are controllable from the start. Marc Cuban, a well-known public figure in the United States and big investor, warns that blindly using data, without considering its faults and errors in for instance high frequency trade, can be the cause of entire markets and economies crashing and resulting in a financial crisis (2014).

The academic world and early big data professionals have to stay on top of technology before it can be implemented into the business. Universities and schools have to develop these skills and prepare a society to work and utilize effectively more sophisticated tools than ever.

References

Accenture, 2014. Retirement Agencies: get More Than a Toe in the Big Data Pool [Online] Available at <http://www.accenture.com/us-en/Pages/insight-retirement-agencies-get-more-toe-big-data-pool-infographic.aspx?c=psv_pubsvccsmq3fb_00012&n=smc_0314#sf23543968> Accessed 23.04.2014

Apache Hadoop.org, 2014. What Is Apache Hadoop? [Online] Available at <http://www.hadoop.apache.org> [Accessed 22.04.2014]

BBC, 2014. Can big data crunching help feed the world? [Online] Available at <http://www.bbc.com/news/business-26424338> [Accessed 18.04.2014]

Brown, P., 2007. Career coach - decision-making. [Print] Pulse

Boyd, Crawford, 2012. Critical Questions for Big Data. [Online] Available at <http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878#.U1frG-PlmN8E> [Accessed 23.04.2014]

Bensinger, G., 2014. Amazon Wants to Ship Your Package Before You Buy It. *Wall Street Journal*. [Online] Available at < http://blogs.wsj.com/digits/2014/01/17/amazon-wants-to-ship-your-package-before-you-buy-it/> [Accessed 20.04.2014]

Brynjolfsson, E., 2013a. The key to growth? Race with the machines (Ted Talk) [Video] Available at <http://www.ted.com/talks/erik_brynjolfsson_the_key_to_growth_race_em_with_em_the_machines> [Accessed 23.04.2014]

Brynjolfsson, E., McAfee, A., 2012. Big Data: The Management Revolution [PDF] Available at <http://hbr.org/2012/10/big-data-the-management-revolution/ar> [Accessed 10.03.2014]

Brynjolfsson, E., Hitt, L., Kim, H., 2011. Strength in Numbers: How Does Data-Driven Decision-making Affect Firm Performance? [PDF] Available at <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486> [Accessed 20.04.2014]

Brynjolfsson, E., McAfee, A., 2012. Race Against The Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy. Digital Frontier Press

Brynjolfsson, E., McAfee, A., 2014. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. 1st Ed. W. W. Norton & Company

Cuban, Marc. 2014. We Have No Idea How Bad It Can Be. CNBC [Interview] Available at <http://www.cnbc.com/id/101539820> [Accessed 20.04.2014]

Cukier, K., Mayer-Schoenberger, V., 2013. Big Data: A Revolution That Will Transform How We Live, Work, and Think. 1st Ed. Eamon Dolan/Houghton Mifflin Harcourt

Davenport, T., 2013. Data Scientist: The Sexiest Job of the 21st Century [Online] Available at <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1> [Accessed 20.03.2014]

DHL, 2013. DHL Thermonet [Brochure-Print] DHL Global Forwarding

Economist, 2012. Driving big data [Online] Available at <http://www.economist.com/whichmba/driving-big-data> [Accessed 10.03.2014]

Evans, J., 2012. Business Analytics: The Next Frontier for Decision Sciences [Online] Available at <http://www.decisionsciences.org/decisionline/Vol43/43_2/dsi-dl43_2_feature.asp> [Accessed 23.04.2014]

Fawcett, T., Provost, F., 2013. Data Science and its Relationship to Big data and Data-Driven Decision making. *Big Data Journal* [Online] Available at <http://online.liebertpub.com/doi/full/10.1089/big.2013.1508> [Accessed 20.04.2014]

Gartner, 2013. Gartner Big data Survey [Online] Available at <http://www.gartner.com/newsroom/id/2593815> [Accessed 14.04.2014]

Ferguson, R.B., 2013. Two legal experts call out the dangers of sophisticated data analytics, left unchecked. *MIT Sloan Management Review* [Online] Available at

<http://sloanreview.mit.edu/article/three-paradoxes-of-big-data/> [Accessed 12.04.2014]

Forbes, 2013. Sears Competes On Big Data and Loyalty Programs [Online] Available at <http://www.forbes.com/sites/tomgroenfeldt/2012/05/02/sears-competes-on-big-data-and-loyalty-programs/> [Accessed 19.04.2014]

Forbes, 2009. Why Data Quality Matters. *Forbes Magazine* [Online] Available at <http://www.forbes.com/2009/08/31/software-engineers-enterprise-technology-cio-network-data.html> [Accessed 28.12.2013]

Google, 2014. Google trends [Online] Available at <http://www.google.com/trends/explore#q=big%20data> [Accessed 23.04.2014]

Haselden, K., Wolter, R., 2006. The What, Why, and How of Master data Management. Microsoft Corporation [Online] <http://msdn.microsoft.com/en-us/library/bb190163.aspx> [Accessed 23.04.2014]

HBR Magazine, 2013 [PDF] Available at <http://hbr.org/magazine> [Accessed 15.04.2014]

IBM, 2014. IBM Big Data [Online] Available at <http://www-01.ibm.com/software/data/bigdata/> [Accessed 15.04.2014]

IBM, 2014. Data Quality [Online] Available at < http://www-01.ibm.com/software/data/quality/> [Accessed 20.04.2014]

IBM, 2013. Big Data & Analytics for Healthcare [Video] Available at <http://www.youtube.com/watch?v=wOwept5WlWM> [Accessed 19.04.2014]

International Association for Information and Data Quality, 2013. Glossary [Online] Available at <http://iaidq.org/main/glossary.shtml> [Accessed 20.03.2014]

Johnson, Whittington and Scholes, 2012. The Cultural Web. [Online] Available at <http://www.mindtools.com/pages/article/newSTR_90.htm> [Accessed 20.04.2014]

King, J., Richards, N., 2013. Three Paradoxes of Big Data. *Stanford Law Review* [Online] Available at <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2325537> [Accessed 10.04.2014]

Lohr, S. 2013 Big Data Is Opening Doors, but Maybe Too Many. *New York Times* [Online] Available at <http://www.nytimes.com/2013/03/24/technology/big-data-and-a-renewed-debate-over-privacy.html?pagewanted=all> [Accessed 23.04.2014]

Manyika, J. et al., 2012. Why Big Data Is The New Competitive Advantage. *Ivey Business Journal* [Online] Available at <http://iveybusinessjournal.com/topics/strategy/why-big-data-is-the-new-competitive-advantage#.U1fWyvlmN8F> [Accessed 20.04.2014]

McAfee, A., 2013. Big Data: The Management Revolution [Video] Available at <http://video.mit.edu/watch/big-data-the-management-revolution-13683/> [Accessed 05.02.2014]

McKinsey, 2011. Big data: The next frontier for innovation, competition and productivity [PDF] Available at <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation> [Accessed 23.04.2014]

Needham, J., 2013. Disruptive Possibilities: How Big Data Changes Everything. 1st Ed. O'Reilly Media

Osty, F., Uhalde, M., 1999. The Social Trajectories of Modernizing French Companies: 299-313.

Payandeh, F., 2013. BI vs. Big Data vs. Data Analytics By Example [Online] Available at <http://bigdatastudio.com/2013/08/24/bi-vs-big-data-vs-data-analytics-by-xample/> [Accessed 23.04.2014]

Peterson, A., 2014. Will the White House's big data privacy initiative distract from the NSA debate?. *The Washington Post*. Available at <http://www.washingtonpost.com/blogs/the-switch/wp/2014/02/10/will-the-white-houses-big-data-privacy-initiative-distract-from-the-nsa-debate/> [Accessed 23.04.2014]

Procter & Gamble, 2013. P&G Business Sphere [Online] Available at <http://www.pg.com> [Accessed 15.02.2014]

Royal Dutch Shell, 2013. Shell Scenarios [Online] Available at <http://www.shell.com/global/future-energy/scenarios.html> [Accessed 24.04.2014]

SAP, 2014. Big data: SAP [Online] Available at <http://www.sapbigdata.com/> [Accessed 18.04.2014]

SAS, 2014. Data Visualization [Online] Available at <http://www.sas.com/en_us/insights/big-data/data-visualization.html> [Accessed 23.04.2014]

Sicular, S., 2013. Gartner's big Data definition Consists of Three Parts, Not to Be Confused with Three "V"s. *Forbes Magazine*. [Online] Available at <http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/> [Accessed 10.04.2014]

Silverman, R., 2013. Tracking Sensors Invade the Workplace. Devices on Workers, Furniture Offer Clues for Boosting Productivity. [Online] Available at <http://online.wsj.com/news/articles/SB10001424127887324034804578344303429080678> [Accessed 23.04.2014]

SITA, 2013. Baggage Report 2013. [Online] Available at <http://www.sita.aero/surveys-reports/industry-surveys-reports/baggage-report-2013> [Accessed 13.04.2014]

Spiegel, et al., 2013. Method and system for anticipatory package shipping [Online] Available at <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=1&f=G&l=50&co1=AND&d=PTXT&s1=%22anticipatory+shipping%22&OS=%22anticipatory+shipping%22&RS=%22anticipatory+shipping%22> [Accessed 23.04.2014]

The Institute for Policy Innovation, 2007. The True Cost of Sound Recording Piracy to the U.S. Economy [Online] Available at <http://www.ipi.org/ipi_issues/detail/the-true-cost-of-sound-recording-piracy-to-the-us-economy> [Accessed 23.04.2014]

United Nations, 2013. Global issues. Agriculture [Online] Available at <http://www.un.org/en/globalissues/agriculture/> [Accessed 20/04.2013]

Abbreviations

HiPPO:
HiPPO stand for Highest Paid Person's opinion and describes the tendency in many corporations to base decisions on the experience of a senior person with extensive experience. This tendency is contradictory to the new generation of managers that often base their decision on data, rather than experience

RFID:
RFID is a technology based on non-contact radio frequency data transmission. This technology is often used to identify items electronically, e.g. in a warehouse. This technology had especially great significance for logistics and inventory management as it had a big impact on operations, similar to the development of the bar code.

Interview questions with professionals in big data

- Information about the interviewee's company:
    a. Industry
    b. Company size
    c. Market share

- Information about the interviewee:
    a. Role and department
    b. Educational background

- How has big data, in your opinion, influenced your duties and job description?
- In your opinion, what significance has big data to your organization?
    a. Which department deals with big data issues?
    b. Does your company follow a big data strategy?

- In your professional opinion, what are the main benefits of implementing big data into the business?

- Comparing big data to the internet, which technology will shape the business more?

- What challenges do you see linked to big data?
    a. Privacy?
    b. Security?
    c. Corporate Culture?
- What is your personal opinion, in general, about big data? What would you like to stress?