LAPPEENRANTA UNIVERSITY OF TECHNOLOGY

LUT School of Business and Management

Master's Degree Programme in Supply Management

Kaisa Rytilahti

EXPLOITING BIG DATA IN A RISK REVIEW RELATED TO SUPPLIER SELECTION

Master's thesis, 2019

1st Supervisor: Professor Veli Matti Virolainen

2nd Supervisor: Associate Professor Katrina Lintukangas

ABSTRACT

Author:	Rytilahti, Kaisa
Title:	Exploiting Big Data in a Risk Review Related to Supplier Selection
Faculty:	LUT School of Business and Management
Major:	Supply Management
Year:	2019
Master's thesis:	Lappeenranta University of Technology, 107 pages, 11 figures, 1 table, 1 appendix
Examiners:	Professor Veli Matti Virolainen Associate Professor Katrina Lintukangas
Keywords:	Big Data, Big Data Analytics, Supplier Selection Process, Supply Risk, Supply Chain Risk Management

Big data is the new hot trend in the business world. Lately big data analytics has been taking massive leaps as a potential and applicable solution to almost every operational challenge company decision-makers are facing nowadays. This research strives to integrate the concepts of big data, supply risk management process and supplier selection process together. The purpose of this research is to find out how big data or big data analytics can be utilized in the processes of supplier risk management and supplier selection to make them more efficient to prevent risky suppliers from even entering the supplier base of companies in the first place. Risk management process and big data analytics are in a sense tools for effective supplier selection. This study also aims to describe the challenges and requirements related to utilizing big data in companies' business operations and to describe the supplier selection process, risk management process in supplier selection and big data to find how they can be integrated with using different methods and technologies to select suppliers. The empirical part of this research is conducted as a qualitative case study investigating two case companies from the same industry. The data for the empirical part was collected with semi-structured and structured interviews. The results suggest that it still requires a lot from companies to start big data related operations and that not all of them are necessarily ready to integrate big data analytics into their business operations. However, at the end this study presents a framework or suggestion how companies could prepare themselves for big data utilization or how they could start to exploit big data in their business processes. Companies aiming to utilize big data should always carefully plan and investigate the execution beforehand.

TIIVISTELMÄ

Tekijä:	Rytilahti, Kaisa
Tutkielman nimi:	Big Datan Hyödyntäminen Toimittajanvalintaan Liittyvässä Riskitarkastelussa
Tiedekunta:	Kauppatieteellinen tiedekunta
Pääaine:	Supply Management
Vuosi:	2019
Pro gradu –tutkielma:	Lappeenrannan teknillinen yliopisto, 107 sivua, 11 kuvaa, 1 taulukko, 1 liite
Tarkastajat:	Professori Veli Matti Virolainen Tutkijaopettaja Katrina Lintukangas
Hakusanat:	Big Data, Big Data Analytiikka, Toimittajanvalintaprosessi, Hankintariski, Toimitusketjun Riskienhallinta

Big data on uusi kuuma trendi liike-elämässä. Viime aikoina big datan asema on ottanut massiivisia harppauksia käyttökelpoisena ratkaisuna lähes kaikkiin operatiivisin haasteisiin, joita yritykset kohtaavat tänä päivänä. Tämä tutkimus pyrkii yhdistämään seuraavat konseptit yhteen: big data, toimitusketjun riskienhallinta ja toimittajanvalintaprosessi. Tutkimuksen tarkoitus on selvittää, kuinka big dataa voidaan hyödyntää riskienhallinta- ja toimittajanvalintaprosesseissa, jotta ne olisivat tehokkaampia. Big datan käyttö toimittajavalinnassa voisi mahdollisesti estää riskialttiiden toimittajien pääsyn yritysten toimittajakantaan. Riskienhallintaprosessi ja big data ovatkin työkaluja joiden avulla toimittajanvalintaprosessista on mahdollista tehdä tehokkaampi. Tämän tutkimuksen päämäärä on myös kuvailla haasteita ja vaatimuksia, jotka liittyvät big datan hyödyntämiseen yritysten liiketoiminnassa. Tutkimuksessa kuvaillaan myös toimittajanvalintaprosessi, toimittajien riskienhallintaprosessi ja big data, jotta voidaan selvittää, miten nämä kolme konseptia voidaan integroida eri menetelmiä ja teknologioita käyttäen toimittajanvalintaan. Tutkimuksen empiirinen osa on toteutettu laadullisena tapaustutkimuksena, jossa keskitytään kahteen yritykseen, jotka toimivat samalla teollisuudenalalla. Empiirisen osan data kerättiin puolistrukturoiduilla ja strukturoiduilla haastatteluilla. Tulokset osoittavat, että se vaatii paljon aloittaa big datan hyödyntäminen yritysten liiketoiminnassa ja että kaikki yritykset eivät välttämättä ole vielä valmiita integroimaan big dataa heidän liiketoimintaansa. Kuitenkin tämän tutkimuksen lopussa esitetään kehys tai ehdotus siitä, miten yritykset voisivat valmistautua big datan käyttöön tai miten he voisivat aloittaa big datan käyttämisen liiketoiminnassaan. Tätä suunnittelevien yritysten tulisi aina suunnitella ja tutkia toteutus huolellisesti etukäteen.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Veli Matti Virolainen and Katrina Lintukangas for their guidance during this thesis project. In addition, I would also like to thank all the interviewees for their time and effort as they enabled me to complete the empirical research of my thesis.
Also, I would like to thank my close ones for all the support they have given me during the whole writing process.

In Helsinki, 28.01.2019
Kaisa Rytilahti

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDICES

1. INTRODUCTION

The most important role big data analytics could have in supply chains is its potential to assist enterprises to manage their supplier risk profile and operating environment (Militaru et al., 2015). Also Milan (2015) states that big data can indeed provide wide opportunities in supply chain management as an efficient tool for supply chain risk management and for measuring supplier performance with a punctuality never seen previously. Actually lately big data analytics has been taking massive leaps in companies from different industries as a potential and applicable solution to almost any operation challenge company decision makers are facing nowadays. (Milan, 2015) This thesis will strive to integrate big data analytics usage and risk review related to supplier selection together; how big data can be utilized to make this process better and more effective, so companies can avoid and control supply chain risks already in the supplier selection phase. It is very important and crucial for supply chain risk management that companies identify the possible risks related to specific suppliers prior to selecting them to their supplier base so those risky suppliers do not even get to enter the companies' supplier base. Big data analytics together with risk management process can enhance the supplier selection process significantly and make it more accurate. This thesis chapter is aiming to give the reader a comprehensive view of the thesis and lead the reader to the topic, explaining why the subject is chosen and justifies why it is important to study.

1.1. Background to the research

One of the most flourishing markets in the next century to come will be big data analytics (Zhong, Newman, Huang and Lan, 2016). It is expected from big data to better the life quality and the "effectiveness" of the world but only if it's potential is exploited and understood. However it is not simple to utilize big data effectively. (Akerkar, 2013, 4) Still the current trend of big data analytics is generating huge potentials and opportunities especially in the supply chain management field (Engel, Sadovskyi, Böhn, Heininger and Krcmar, 2014). Also Wang, Gunasekaran, Ngai and Papadopoulos (2016) state that big data analytics has gained a growing attention especially in the logistics and supply chain management field. This is due to the complex characteristics of big data and the significant part these fields have in enhancing the whole process of doing business. (Wang et al., 2016) Big data is

predicted to enhance the processes and collaboration in supply chains (Waller and Fawcett, 2013).

According to Veldhoen and De Prins (2014) many industries do already utilize big data. Still risk management field is not yet using big data in a large scale despite the fact that big data can better the risk models' predictive force, deliver wider coverage for risk management, enhance risk system effectiveness and response times, and create substantial cost savings in risk management. (Vedhoen and De Prins, 2014) Management of risks is one example of the most substantial challenges but also important issues supply chains are facing in the 21st century (Torres-Ruiz and Ravindran, 2018). According to Fan, Heilig and Voss (2015) especially the risk management of supply chains can hugely benefit from big data analytics and technologies that are excellent for monitoring, collecting and analyzing both supply chain environmental and internal data. (Fan et al., 2015) Big data analysis indeed can be utilized to forecast events to prevent risks from realizing (Engel et al., 2014). A supply chain failure can realize at any stage of the process and that is why proactive and preventive supply chain risk investigation is necessary (Mohtasham, Aziz and Ariffin, 2016).

Firms that rely too much on their suppliers are vulnerable to risks if the supplier is not able to fulfill their demands. Therefore managers of supply chains must constantly reduce the risks from the supply chain. For instance a natural disaster in Asia should not cut the supply in Europe. (Singh, Jain, Mehta, Mitra and Agrawal, 2017) One of the basic and most significant phases in the risk management of supply chains is the supplier selection. It is also very critical for the reduction of risks in the whole supply chain. The supplier failure is one of the biggest risks the supply chains encounter and that is why taking care of proper supplier selection is very important for companies. Risks and uncertainties caused by suppliers need to be investigated and included into the supplier selection decision-making. Companies need to recognize the critical supplier risk factors, so they can choose the reliable suppliers and thus have a flexible supply chain that is capable to react to the risks and uncertainties. It is top importance to carefully evaluate the potential suppliers and assess their performance whether they are contributing uncertainties and risks to the supply chain. (Mohtasham et al., 2016)

Selection of suppliers is a hard task in any business (Patra and Mondal, 2015). Especially the global supplier selection is a problem way more complicated than domestic supplier selection because it requires a lot more analysis. Major risk factors need to be analyzed since the selection between several unknown international

suppliers is very critical supply chain decision for any company. Suppliers need to be compared based on vast criteria and measure base. The process of selecting suppliers is very important since it has an immediate impact on organization's performance. (Chan and Kumar, 2007) Also Chen and Zou (2017) and Patra and Mondal (2016) state that supplier selection and assessment is really a critical and important process that affects the company's revenues and risks. Vedhoen and De Prins (2014) state that the success of risk management will be defined by the utilization and access to big data sources in today's world that is more and more complex and demanding. Big data is the future of the management of risks – the game changer for risk management is big data. (Veldhoen and De Prins, 2014)

1.2. Research questions and research objectives

The main purpose of the thesis is to find out how big data can be exploited in a supplier risk review related to the supplier selection process; how it is possible to utilize big data in the process to make it more effective. The sub-questions of the thesis are used to help in responding to this main research question. The objective of the first and second sub-question is to construct understanding of what is precisely meant by supplier selection process and the supplier risk review and the criteria related to selecting the suppliers. The third sub-question tries to answer what is exactly meant by the big data concept that is a highly complex concept and does not even have one widely agreed and unified definition. The fourth sub-question seeks to answer what companies need to have to be able to utilize big data in their supplier selection risk review process and what are the challenges companies can encounter when using big data in their business operations. The fifth sub-question tries to determine the big data related methods and techniques companies can utilize in supplier selection risk review process. Without fully understanding the background behind big data and its utilization it is impossible for enterprises to start using big data in their supplier selection and supplier risk management processes.

The main research question:

- How big data can be exploited in a risk review related to supplier selection to make the process more effective?

The sub-questions:

- What is the supplier selection process (and the criteria behind the selection)?

- What is the risk review of suppliers in the supplier selection process (and the risk criteria that must be assessed)?
- What is big data?
- What companies need to have to be able to utilize big data and what are the challenges of using big data?
- What big data related methods and techniques are there that can be utilized for supplier selection risk review?

1.3. Conceptual framework and key concepts

This thesis investigates how big data analytics and risk management process can be utilized together in the selection of suppliers and how the concepts can be used to enhance supplier selection process and, in the end, apply additional value to the whole supplier selection process (Figure 1). The main purpose is to find out how risk management process and big data analytics can be used as a sort of tools for supplier selection to ensure the correct suppliers get selected to companies' supplier base. This way the thesis provides a new perspective to supplier selection process. There is a huge potential in the utilization of big data analytics and risk management process together in the supplier selection to enhance and make the process a lot more effective and accurate and so generate added value to the supply risk management and selection processes and give companies competitive advantage in this way.

The core behind the conceptual framework is the basic risk management process, where risks are first identified and analysed, then evaluated and treated accordingly. According to Tummala and Schoenherr (2011) the risk management of supply chain has usually the following steps: risk identification, risk assessment and measurement (analysis and evaluation), and risk mitigation and contingency plans (risk treatment). Also risk control and monitoring via data management systems is happening on the background constantly and for that to succeed constant communication and consultation are also important side functions in the process. (Tummala and Schoenherr, 2011). The capability to recognize suppliers having a higher potential of risks and disruptions is critically important for companies since supplier risks have a massive effect on the supply chain (Trkman and McCormack, 2011). Especially when supply chains are facing risks and uncertainties, it is even more crucial to manage the risks beforehand already in the supplier selection phase (Mavi et al., 2016). Big data can especially be used to enhance company's supplier selection and risk review process (GEP, 2018). Many sources and different authors indicate that usage of big

data analytics generates value to the whole process and makes the supplier selection more accurate.

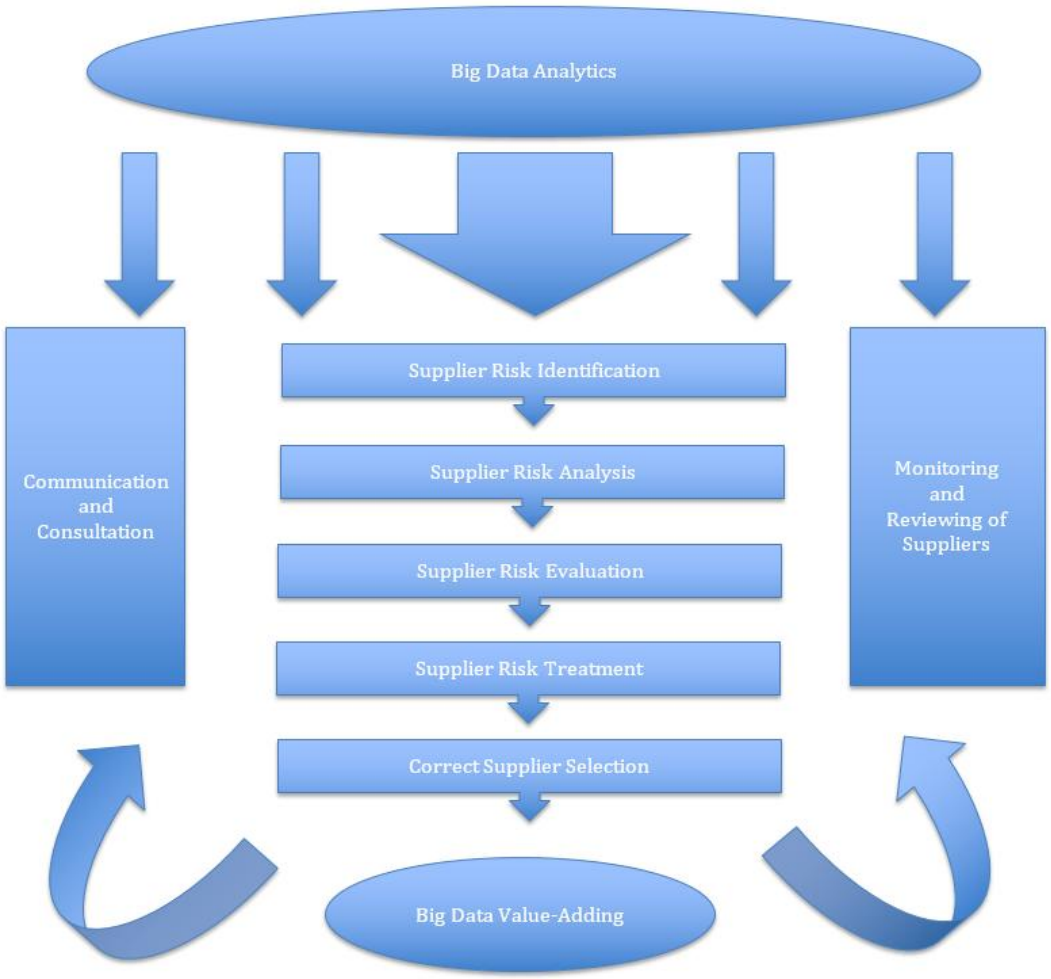

Figure 1. Conceptual framework

Big data is used to refer quantitatively and qualitatively novel types of data but also to revolutionary ways of collecting data, processing and storing it (Engel et al., 2014). According to Militaru et al. (2015) the big data concept is utilized to describe large sets of data that are so massive that it overruns the computer memory. Veldhoen and De Prins (2014) state that big data has certain complex characteristics such as volume, velocity and variety that need novel technological organization approaches and analysis. Originally big data is to the vast data flood measured in exabytes and even more (Zhong et al, 2016). According to Singh, Jain, Mehta, Mitra and Agrawal (2017) big data can be structured or unstructured data. Big data is so huge that it is hard to handle utilizing ordinary database and software technologies. Big data can also be determined as the use of data to manage processes and to make decisions (Meraglim, 2017).

Big data analytics is the usage of novel and progressive analytic technologies on big data such as data mining as a practice of business intelligence. Big data analytics or BDA is analyzing of massive data sets with varied types to make decisions by exposing hidden trends, correlations and patterns, and other knowledge and information that is useful for business to be able to gain business benefits such as operational efficiency and to investigate novel opportunities and markets. Descriptive, predictive and prescriptive analytics are the three categories of big data analytics. (Tiwari et al., 2018)

Supplier selection process and decision is a multi-objective and multi-criteria problem. It has a huge strategic significance on enterprises and the decision nature is often very unstructured and complex. Supplier related criteria, service and product performance criteria and cost criteria are the most traditional selection criteria are. (Kahraman, Cebeci and Ulukan, 2003). The selection of supplier is decision based on multiple qualitative and quantitative factors (Ho, Xu and Dey, 2010). The suppliers are assessed, surveyed and chosen to have a part in the company's supply chain. The process of selection has a huge strategic importance and it has many uncertainties and risks related to it and that is why often many decision makers from multiple departments of the company that have a part in the decision-making. (Sanayei, Mousavi and Yazdankhah, 2010) Basically the process is about choosing the fitting suppliers at the right time, at the correct price, with the correct quality in the correct quantities. It is about making a decision and evaluating suppliers for contract making. (Mavi, Goh and Mavi, 2016). According to Sonmez (2006) the process of supplier selection begins with the identification of the demand for a supplier. The following steps after this are: decision criteria formulation and determination, pre-qualification (selecting only few possible ones from a huge supplier list), final selection of supplier and the monitoring of suppliers (constant assessment and evaluation).

Supply risk is defined by Peck (2006) but also by Mavi et al. (2016) as any occurrence or disrupt that prevents or disturbs product, information and material flow of the final product from the original supplier through the supply chain to the end customer. Wu, Liao, Tseng, Lim, Hu and Tan (2016) say that functional risks and triggering events are synonyms to supply risks. Functional risks are the appearance of an unexpected problem in enterprise's fundamental functions an recognizing triggering events is the basis for recognizing and lessening uncertainties and risks (Klinke and Renn, 2002). According to Borghesi and Gaudenzi (2013) supply chain risks have originally been divided into four groups that are financial, hazard, strategic, and operational risks. Supply chain risks are connected to the supply chain network's strategies, activities and nature. All of these are possible sources of supply chain risks. (Borghesi and

Gaudenzi, 2013, 117-118) Generally risk is defined as losses or harmful consequences. The risks are losses and uncertainty of their amount and occurrence. The risks arise from uncertainty. (Hallikas, Karvonen, Pulkkinen, Virolainen and Tuominen, 2004) There are multiple diverging categorizations of supply risks and there is no unified answer to this either, which is the correct categorization.

Supply chain risk management has the objective to recognize the areas of potential internal and external risks in the supply chain and to conduct needed activities to control the risks. (Torres-Ruiz and Ravindran, 2018) The objective is to recognize the possible supply risk sources and carry out suitable operations to be able to avoid the risks (Narasimhan and Talluri, 2009). According to Mavi et al. (2016) management of supply chain risks is possible with an organized and comprehensive approach between supply chain partners to mitigate supply chain vulnerability. SCRM is used on daily basis but also for exceptional risks to ensure continuity of business proactively. There are also two major joint SCRM customs: information sharing of risks and mechanism for risk sharing. Managers choose the right risk strategies considering multiple attributes such as the risk origin, risk nature and company resources. Different risk types also need different methods to manage them. The SCRM process has five steps: identification of risks, assessment of risks, analysis of risks, risk management procedures; and risk monitoring and evaluating (Lester, 2014, 71).

1.4. Literature review and research gaps

Research of big data popped up in the 1970s but the publications regarding big data and its utilization in various fields have grown exponentially only since 2008 (Addo-Tenkorang and Helo, 2016). All in all big data has not been extensively researched from any specific viewpoint. Big data researches have for instance aimed and focused at investigating opportunities, challenges and trends of big data. Much research has been made for example from the viewpoint that what big data is as a concept to define it more precisely. Still Addo-Tenkorang and Helo (2016) say there is a restricted agreement about big data performance and what is its most value-adding use. There is not even a single accepted definition to big data even though the big data challenges, impact and performance is handled across different sectors.

According to Waller and Fawcett (2013), and Rozados and Tjahjono (2014) there is not much literature or research about the utilization of big data in supply chain management and how it can have an effect on management of supply chains. Also

Sadovkyi et al. (2014) implies that there is overall a huge lack of novel practical and theoretical progress in the field of supply chain management. Wang et al. (2016) state that there is actually a gap between supply chain practices and theory from the supply chain analytics viewpoint. In addition there are also still many difficulties in applying big data solutions in real life (Militaru et al., 2015). However big data in management of supply chain has been investigated in some specific industries such as in service and manufacturing industries like Zhong et al. (2016) in their study. Some researchers have also studied big data from the perspective of Twitter analytics usage in supply chains that is one sort of social media analytics method according to Chae (2015). Still Wang et al. (2016) state that existing research on big data business analytics on supply chain management and logistics have mostly concentrated on analyzing different perspectives and definitions or recognizing possibilities for supply chain education and research. Overall the big data business analytics are in its infancy and there are still researches to be made to investigate business analytics of big data in different contexts of logistics and supply chain management. Also according to Tiwari, Wee and Daryanto (2018) in the current researches there is not really any consensus about the big data performance in supporting supply chain management.

According to Tummala and Schoenherr (2011) overall risk management of supply chains is an emerging but significant research line in the interconnected and dynamic world that we live in. There are not much guidance or conceptual frameworks on the topic. (Tummala and Schoenherr, 2011) From the viewpoint of risk management big data has mainly been investigated in the bank and financial industry. In the light of supply chain risk management, for instance Ratnasingam (2006) has conducted a research aiming at discovering possible features of supply chain uncertainties and risks. However according to Wu et al (2016) this study but also previous researches mishandle the inter-relationships among different supply chain risks. Wu et al. (2016) have researched big data and supply chain risk and uncertainties management from the viewpoint of sustainability to explore some decisive attributes. Chen, Tao, Wang and Chen (2015) investigated in their study fraud risk management at Alibaba that is based on big data.

According to Chan and Kumar (2007) in the past many researches have focused only on domestic supplier selection and thus left many very critical global criteria un-discussed. According to them there is only limited amount of discussion in previous researches about global supplier selection process. In their research the question of global supplier selection is tackled with fuzzy methods techniques. According to Ruhrmann et al. (2014) methods existing for supplier selection are not taking so well

into account the risks, dynamics and forecasting methods. The existing methods are mainly focused on monetary supplier criterion. Also according to Rao, Xiao, Goh, Zheng and Wen (2017) most of the researches in supplier evaluation do not consider the supply chain environment risk factors and has concentrated on commercial criteria such as lead time, quality, and price. Also according to Chen and Zou (2017) much research has been made about the supplier selection but only few researches has investigated the problem from the point of few of risk aversion. Much research from supplier selection has been made with for example fuzzy techniques as mentioned before and for example with Delphi method. According to Foerstl, Reuter, Hartmann and Blome (2010) there is not much research either on how companies make a decision about what are their suppliers' risks or how suppliers construct their supply risk management to make sure they are not exposed to the risks caused by suppliers. According to Patra and Mondal (2015) and Rao et al. (2017) supplier selection is becoming on of the most researched and hottest supply chain questions.

According to Trkman and McCormack (2011) even risk management of supply chains is rather novel concept and thus it is currently a bit chaotic and disorganized. There are multiple different risks and methodologies classifications and usually they concentrate on the forecasting of disruptive events such as natural disaster, terrorist attack and bankruptcy and do not investigate deeply the root causes behind the uncertainties and risks. Constant changes due to a turbulent environment (technology changes, changes in customer tastes or supplier priorities) are not that much investigated. This approach to risk management is not taking into account the fact that environmental, market and technology turbulence in the supplier's specific market have a high impact on potential disruptions, relationship between supplier attributes and supply chain performance. Further because multiple suppliers do business in different environments and markets, their turbulence vary and thus the forces having an impact on supplier are also different. That is why all strategies to manage supplier risks do not work for every market and supplier. For example in a market where technology is constantly rapidly changing, company cannot mitigate risks by having a buffer stock. That is why it is important to have a comprehensive approach in SCRM and take into account supplier-associated turbulence and multiple uncertainty sources because of supplier features such as performance, structure and strategy. Companies also need to note that there is no one right way to manage supply chain risks instead risk management is quire firm-specific.

1.5. Delimitations

This research is limited to handle the concepts of big data analytics, supplier selection and supplier risk management together to be able to answer the research questions of this study. The thesis is concentrated on how big data can be utilized through risk management process in risk review related to supplier selection to make the process of supplier selection more precise and effective. There are many supplier risks and other supplier selection criteria related to supplier selection that need to be investigated but because here the limitation is already the supplier selection process from the supplier risk management review point of view and how big data and risk management process can be used together in supplier selection, there is no need to limit the subject further to some specific risk type such as external environmental risks or internal supplier specific risks. Also, there is no need to limit the supplier selection process or supplier risk management process any further as the big data concept already limits the subject. Further the research is done from the companies point-of-view. Overall the research questions and objectives limit the theoretical part of this thesis.

The theory part of this thesis focuses on general level on all kind of companies in all kind of industries. Because the purpose of the thesis is to uncover how big data can enhance the risk review in supplier selection and what big data actually is as a concept, there is no need to limit the research into some specific company, industry or country or into one specific risk type that need to be investigated when selecting suppliers and examining their risks. Also, there is not that much of a research on the topic of big data in the supply chain management side, so it would have been very difficult to get enough information for the theoretical part if it would have been limited to concern for example only one specific industry. However, in the empirical part the both case companies are from mining industry so that naturally limits the empirical findings part. Because a more comprehensive picture of the big data effect on business processes is sought, a qualitative case study approach in the empirical part is appropriate.

1.6. Research methodology and data collection plan

The theory part of this thesis is constructed of scientific articles and other literature such as books and previously made researches but also some Internet news and articles. The topics of the sources are connected to big data, big data analytics, supply chain management, supply risk management and supplier selection process and their risk review related to the selection process. Those key concepts were used when

searching for literature for this thesis. In this way it has been made possible to construct a theory part for the thesis that is meant to build an extensive basis for the empirical part that is presented in this thesis right after the theory part.

Because of the design of the research problem and questions of this thesis, the research method for the empirical part of this thesis is a qualitative research. Many phenomena that are related to management such as risk management, organizations and markets require from the research that the phenomena investigated is approached with qualitative research to gain more understanding (Koskinen, Alasuutari and Peltonen, 2005, 15). The aim of qualitative research is to describe, explain and understand (Gibbs, 2007, 94). Qualitative research is also suitable for the topic of this thesis in sense that it offers a way to withdraw from theoretical and conceptual customs that guide mainstream researches. A carefully done qualitative research is enough as it is without quantitative research. Further qualitative case study research is suitable for this research as it is used in situations in which the subject nature is in a need of deeper understanding. (Koskinen et al., 2005, 23-25)

The empirical part's case studies are made from qualitative material collected by interviewing employees from two companies that are interested in big data and have a supply chain department. Case study is suitable for the topic of this thesis because by using case study it is possible to obtain understanding of complexity and to get specificity to the thesis topic but also because with case studies it is possible to gain a comprehensive picture of the companies (Koskinen et al, 2005, 156). Case studies aim to understand the research topic more profoundly (Metsämuuronen, 2005, 222). The case studies are done as semi-structured interviews alias theme interview by interviewing people that are in touch with big data utilization in case companies or work within supply chain management. In theme interview the interview is usually implemented with questions that are made by the interviewer that the interviewee can answer freely in own words. Theme interview is used in this thesis because when conducted carefully it is very efficient way of making qualitative research. This is because the interview can be guided without controlling it entirely. (Koskinen et al., 2005, 104-5) However the interviews conducted via email are structured interviews.

1.7. Structure of the thesis

This thesis is constructed of two main parts that are the theory part and the empirical part. The theory part supports the empirical part that follows the theory part. Both parts

seek to answer to main research question of the thesis but also to the sub-questions that are needed to better answer to the main research question. After the theory part's introduction chapter, the thesis deals with important theory that is related to the topic of this thesis. The theoretical part is divided into four separate chapters according to the larger themes of the thesis. The first three themes are big data, supplier selection process, and risk management related to supplier selection. The last big theme is how big data can be integrated using different big data analytics methods and technologies to companies' supplier selection risk review and at the same time the chapter strives to integrate the three earlier presented themes. The empirical part of this thesis follows the theory part. In the empirical part the thesis investigates as a qualitative case study two case companies that are interested in using big data in their business processes and have a supply chain department. Before presenting the empirical findings of this thesis the research methodology and data are explained in more detail. The last chapter of this thesis consists of discussion and conclusions of the thesis. Also a suggestion or a framework for starting to use big data analytics in companies' supplier selection risk review is presented and the benefits that big data analytics utilization in supplier selection risk review generates. The last chapter also discusses theoretical and managerial contributions of this thesis but also the limitations and future research suggestions stemming from this thesis are handled.

2. BIG DATA

Big data as a concept does not have one unified definition. Bikakis (2018) state that the big data era has generated the availability of huge volumes of vast data sets that are heterogeneous, dynamic and noisy by the nature with high volatility and variety. In turn according to Tiwari et al. (2018) big data is complex or huge data sets that have a range of exabyte and even more. Addo-Tenkorang and Helo (2016) define big data as fast and constantly growing data amount that comes from multiple different sources that progressively cause enterprises various challenges and complex problems related to analysis, storage, valuable-use and storage problems. Big data is also the datasets that is impossible to acquire, perceive, store, manage and analyze by software or hardware systems and legacy IT in a reasonable time. (Addo-Tenkorang and Helo, 2016) In other words big data is data whose data representation; data volume or acquisition speed prevents the use of classical management methods of database to perform efficient analysis (Mayer-Schönberger and Cukier, 2013). For companies it is important to do constant diversification of big data content. It is important to recognize the right and essential data and to be able to react to the processed information rapidly (Salo, 2014, 6).

Big data is constantly cumulated and is coming from various sources and it can be unstructured or structured (Militaru et al., 2015). All kind of sensors are constantly streaming data throughout the company (Akbay, 2015). An enormous proportion of this data is generated in the supply chain networks' appliances such as smartphones, computer systems, computerized appliances and embedded sensors. (Addo-Tenkorang and Helo, 2016) Big data is something that can be seized, informed, maintained, analyzed and aggregated if that is done properly with the right technologies. The parallel computing methods such as cloud computing can help in this by making the analyzing and acquisition of big data more effective. In some sense big data has expanded the technological capability scope to manage, store, interpret, visualize and process huge data amounts (Kaisler, Armour, Espinosa and Money, 2013). The worldwide digital technologies usage has generated the big data business analytics or BDBA emergence (Chen et al., 2012) that contains big data and business analytics (Wang et al., 2016). Big data is present in nearly every industry and provides companies new abilities to get insights from their business operations (Blau and Gobble, 2015).

According to Addo-Tenkorang and Helo (2016) big data is a trending new enterprise platform or system that presents features for analyzing, acquiring and storing huge amounts of data from multiple sources to gain value. Akbay (2015) state that if data is properly used it can generate remarkable business advantages for companies. Big data makes it possible to connect business rules to the data streams. This produces opportunity to inform systems and people in real time. (Akbay, 2015) Many companies' industrial supply chain management experts and stakeholders forecast that big data will have an upbeat effect on their activities and operations making it possible to make more informed and strategic data-oriented decisions. (Addo-Tenkorang and Helo, 2016) Also Militaru et al. (2015) state that big data generates novel growth opportunities for enterprises from supply chain by having gathering and analyzing the data of services and products, suppliers and buyers, customer intent to buy, and performance. If companies invest to generate their supply chain's big data abilities, they improve their long-term competitive advantage. (Militaru et al., 2015)

2.1. Characteristics

Big data has certain characteristics that distinguish it from other data (Manyika, 2011). The special characteristics and features of big data can advise the enterprise risk management analytics and scenarios, generate better profits and growth, and advance the growth of company that is aware of its risks to avoid loss events in the long term. However the characteristics of big data also cause a lot of challenges for companies. Only if enterprises see these challenging big data characteristics as possibilities and understand them companies can create real business value. (IBM, 2014) Tiwanti et al. (2018) state that the main characteristics of big data are included in the "5V" concept that is constructed of volume, velocity, variety, and veracity but also value. Emani et al. (2015) impose that to efficiently deal with big data enterprises need to generate value against variety, veracity and volume characteristics of data while it is still in motion (velocity) and not after it is in rest as then it is too late. Enterprises need to jointly handle big data while taking all of its characteristics into account. Companies have to have a comprehensive picture of the big data characteristics to take the most out of the opportunities and potential of big data. (Emani et al., 2015) Sivarajah et al. (2017) has added to the most common concept of 5V also variability and visualization as big data characteristics.

2.1.1. Volume

Volume is the huge data sets consisting of terabytes, petabytes and zetabytes of data. Already the pure volume and huge scale of data is a massive challenge for big data processing. (Sivarajah et al., 2017) According to Raguseo (2017) volume is the production and gathering of massive amounts of data where data scale is increasingly high. Volume is referred to the fact that data size in the world is growing exponentially all the time (Salo, 2013, 21). Tremendous amount of big data is created staggeringly every moment within supply chains worldwide (Zhong et al., 2016). According to Philip Chen and Zhang (2014) big data volume refers to the data size so huge it is almost impossible to comprehend. Volume is basically the big data quantity.

Big data can be characterised and described in multiple ways and one is to divide it into two parts: static and flowing data (Figure 2). A metaphor from nature can be used to describe this breakdown. Data in data warehouse is ocean and constantly flowing and moving data is river. An example of flowing data is for example the amount of video material that can be massive thanks to multiple high quality cameras – so massive that the present data warehouses cannot record and save them. (Salo, 2013, 23-24) Flowing data is produced by for instance sensors, transmission networks, machines and devices, cameras, cloud services, media services, and transactional systems. (Salo, 2013, 60) According to Emani et al. (2015) the most important appeal of big data analytics is the capability to process huge volumes of data. That is why enterprises are storing huge volumes of data of various types depending on their need. This stored data is in a sense resting.

Figure 2. Static and flowing data (Salo, 2013, 23-24)

2.1.2. Velocity

Velocity is the high inflow rate of data whose structure is non-homogenous. It is a challenge to manage the high flood rate of data that is non-homogenous that leads to

either updating the old existing data or creating new data. (Sivarajah et al., 2017) According to Raguseo (2018) velocity is used to refer to the timeliness by which data is being generated, gathered and analyzed. It is the accelerating pace at which data is being fed to the information system and at which speed the data needs to get from there to use (Salo, 2013, 21). To put it shortly velocity is the speed of outgoing and incoming data (Philip Chen and Zhang, 2014). It is the data in motion (Emani et al., 2015).

As velocity is the speed of new data creation it creates the need for data analysis in real-time in a timely manner to get value (Engel et al., 2016). Velocity is also very critical for enterprises as it defines the lag time or latency between the time that data is created and when it is usable for enterprise decision-making (IBM, 2014). The velocity of dealing with massive data sets from supply chain is very important since decisions that are data driven need to be made fast. The velocity characteristic hugely depends on the data transferring reliability, data collection speed, algorithms and models for decision-making, efficient data storage, and excavation speed discovering useful knowledge. (Zhong et al, 2016) The barrier or challenge for companies with this big data characteristic is that because the data is generated nowadays at such vast pace it exceeds the power of many systems and technologies to recognize at the right time possible risk happenings for action and analysis (IBM, 2014).

2.1.3. Variety

Variety is used to refer the varying data types that can be unstructured or structured and coming from many different sources such as text, image, multimedia, audio, video etc. The challenge is to handle this data that has very dissimilar and diverse heterogeneous forms (Raguseo, 2018). Also Philip Chen and Zhang (2014) define variety as the types and sources of data that are very different from each other. (Philip Chen and Zhang, 2014) Novel data types can proliferate from many different sensors that are being utilized in retailer shops, manufacturing sites, facilitated houses and highways, trucks and mobile phones. It demands a more universal and complicated makeup language to integrate such versatile data into standard formats. (Engel et al., 2016) This big data characteristic is a challenge for enterprises because they usually tend to rely greatly on internal source of data and ignore the external data. (IBM, 2014)

Very structured data can originate from relational databases. On the other hand semi-structured data comes from web logs, social media feeds, e-mail, or its raw feed from

sensors. Unstructured data originates from still images, video, clicks and audio. (Emani et al., 2015) According to Huda et al. (2018) structured data originates from companies resources data. Salo (2013, 22) states that this diversification of data into structured and unstructured data does not do justice to the diversity of data. Actually it is preferable to talk about continuum in which there fits a lot of intermediates between the two extreme forms that can be called semi-structured data (Figure 3). An example of semi-structured data is video material or pictures that are equipped with keywords. The video itself is unstructured data but the keywords such as the camera name are structured data. However most of the data is still unstructured and thus kind of useless.

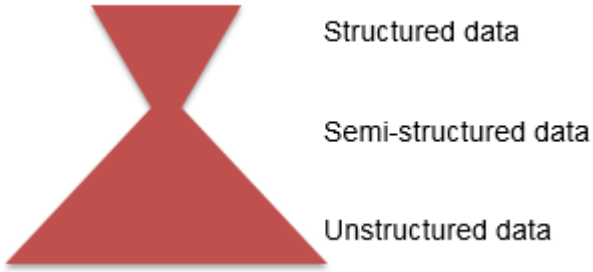

Figure 3. Structured, semi-structured and unstructured data (Salo, 2013, 22)

2.1.4. Veracity

Veracity is the increasingly complex data system, but also inconsistency, imprecision and anonymities that exist in big data. This characteristic is all about data quality and about understanding the data because the data often contain essential inconsistencies. (Sivarajah et al., 2017) Veracity is the trustworthiness and reliability but also the messiness of the data (Chen and Zhang, 2014). According to Emani et al. (2015) veracity is the fact or truth that lays in the big data and the uncertainty can originate for instance from model approximations, inconsistencies, deception, ambiguities, duplication, fraud, spam, latency and incompleteness. According to Engel et al. (2016) veracity means that enterprises need to blend innovative skills and technology to deploy the characteristic V's of big data to transform the data into business information that is useful.

According to Zhong et al. (2016) there is a lot of bad data such as imprecise attributes and noise in supply chain big data. This bad data should be verified to be able to pick the good and useful data that companies can exploit. The verification process should be made under certain security levels and authorities and it should be developed and

designed as automatic tool to verify the compliance and quality issues of data. It might weigh various situations that might be so complicated that it is hard to even address them. (Zhong et al., 2016) Enterprises need to control the uncertainty of specific data types such as data from social networking, sentiment analysis and physical security access data since these data types have very precious information that can recognize possibilities for risks. Because of the veracity dimension enterprises do not for example trust the information they are using in decision-making and are unsure of how much the data they use is inaccurate since the costs of poor data quality are huge for companies. (IBM, 2014)

2.1.5. Value

Value refers to trying to extract value and knowledge from massive amounts of unstructured and structured data without losses. Researchers of big data believe value is an essential feature of big data because without any value and benefit big data is useless. Somewhere in that data there is information that is valuable called high-valued or golden data. However it is hard to extract value from data cost effectively. (Sivarajah et al., 2017) Value is the worth of hidden insights inside big data (Chen and Zhang, 2014). The value characteristic is the fundamental purpose and outcome of using big data technology. The whole point of big data technology is to economically get value from vast amounts of various data types by making high-velocity capture, finding and analysis possible. (Emani et al., 2015)

The value can be divided in two groups: analytical use (support and replacement of human decisions, populations segmentation to customize actions, needs discovery) and enabling novel services, products and business models. (Emani et al., 2015) However value of big data is hard to evaluate in supply chain management context. Getting value from big data is difficult due to the challenges caused by other big data characteristics. Same time it is hard to investigate the impacts and the benefits that big data gives companies, processes and insights within supply chain management. (Militaru et al., 2015)

2.1.6. Visualization and variability

Sivarajah et al. (2017) has added two big data characteristics to traditional 5 V’s described earlier: visualization and variability. Visualization refers to presenting the

data in a readable manner. Visualizing data means presenting the key knowledge and information more efficiently and instinctively by utilizing different visual formats such as graphical or pictorial layout. Variability characteristic on the other hand is used to describe data whose meaning is changing. constantly. Variability is usually confused with variety even though it is also an essential feature of big data. For example Facebook and Google produce various data types. If one of these varying data types is used for analysing, data offers every time a meaning that is different. This is the data variability whose meaning is changing constantly and fast. Variability is also used to refer to sentiment analysis. For instance a word can have several varying meanings in a Tweet. To be able to perform a proper sentiment analysis, the used algorithms have to be capable to comprehend the context in which the word is used. However this is very difficult task. (Sivarajah et al., 2017) Also according to Emani et al. (2015) variability highlights the language meanings' and communication protocols' variability or semantics.

2.2. Sources of big data

Data amount in the world is constantly growing (Figure 4). These days all manufacturing and service sectors are encountering a tsunami of data (Zhong et al., 2016). Especially the global supply chain management industry is having a huge and increasing big data information amount that is flooding from different sources in real time such as sensor networks, digital machines, and mobile equipment. Also geospatial devices generate big data and all of these sources have the ability to substantially advance the accuracy of supply chain management process if the data is being used well. (Zhong et al., 2016) Historically humans have produced the data but now also lifeless objects are generating more data every year than humans have ever produced. Machine data is making data all the time in growing variety, volume and velocity. (Akbay, 2015) Also the whole enterprises are possible sources of data (Blau and Gobble, 2015).

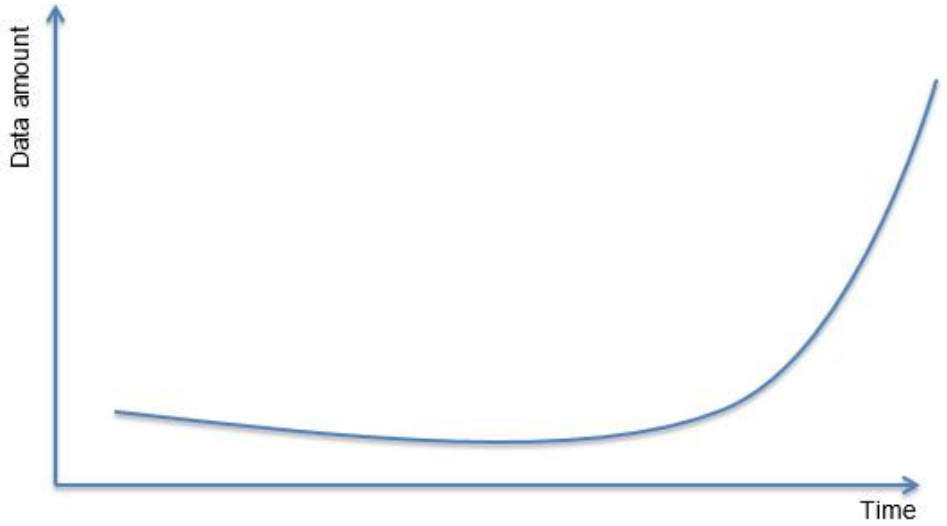

Figure 4. Graph of data amount growth over time

There are multiple different big data sources (Figure 5). User generated content is for example messages such as emails, tweets, texts and blogs. Transactional data are for instance business transactions and web logs. Scientific data is data that is stemming from experiments that are data-intensive such as healthcare and genome data. Web data is for instance sensor data readings and images posted on social media. (Sivarajah et al., 2017) It is possible to highlight the importance of big data with the point that data is generated extensively every day from multiple sources in many forms in unparalleled volume, velocity and variety. In only one minute over than 98 000 tweets are written, 695 000 Facebook status updates are posted, 11 million messages and over 169 million e-mails are sent, 685 445 Google searches are done, over 1820 TB of data is produced and 217 new mobile web users every minute. (Raguseo, 2018) All of this data is obtainable nearly immediately and thus is producing chances for analysis almost in real-time. (Veldhoen and De Prins, 2014)

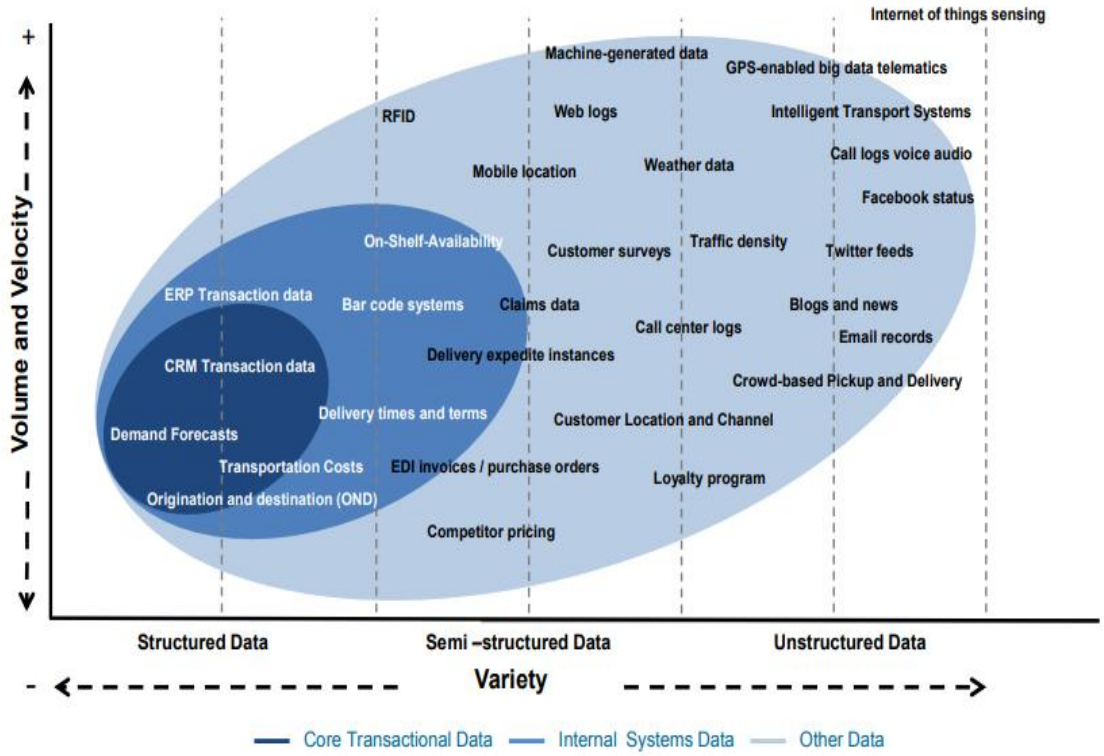

Figure 5. Data sources classification in supply chain (Rozados et al., 2014)

Enterprises are mainly interested in two big data types. The first one is created by humans for example through web tools such as emails, social networks and cookies. Secondly companies are seeking to combine data created sources that are connected. The Internet of human being and the Internet of things have to become a big data mix that needs to be focused to be able to plan and operate predictively. (Emani et al., 2015) Big data is a combination of various data types. The most important big data sources are Internet of Things (IoT), multimedia, self-quantified and data from various social media platforms. (Yaqoob et al., 2016) The data can also come for example from Global Positioning System (GPS), call centers, radio-frequency identification (RFID), point-of-sale systems (POS) and even Facebook. (Militaru et al., 2015) In addition for example manufacturing sector has a vast volume of data stemming from digital machines, electronic devices and sensors that are utilized in shop floors, factories and production lines. (Zhong et al., 2015)

3. SUPPLIER SELECTION PROCESS

Efficient supply chain management requires central role for the selection process of suppliers (Rao, Xiao, Goh, Zheng and Wen, 2017). According to Amid, Ghodspour and O'Brien (2009) as companies' external partners the suppliers have a massive effect on the performance and competitive advantage of the whole supply chain and that is why supplier selection is the most critical decision to be made in a supply chain. (Amid et al., 2009) Good supplier selection reduces costs significantly (Sanayei, Mousavi and Yazdankhah, 2010). According to Guneri, Yucel and Ayyildiz (2009) but also according to Ruhrmann et al. (2014) supplier selection and especially global supplier selection is a choice with high degree of uncertainties, risks and fuzziness. Further global supplier selection is much riskier than domestic supplier selection because highly competitive environment has made firms highly dependent on other companies in the supply chain (Chan and Kumar, 2007).

According to Ho, Xu and Dey (2010) but also according Ng (2007) and Li, Yamagucki and Nagai (2007) companies have to make the selection of suppliers based on multi-criteria to make a good and comprehensive selection decision. (Ho, Xu and Dey, 2010) The decision is also a multi-objective decision (Zeydan, Colpan and Cobanoglu, 2011). This is why decision-makers see supplier selection as a complicated decision to decide with many quantitative and qualitative factors that need to be considered. Firms have to select those suppliers that can increase company's supply chain competitiveness and not decrease it. (Mohtasham et al., 2016) Poor decisions in supplier selection could directly lead to critically very bad consequences for companies such as massive delivery delays and bad customer service. (Chan and Kumar, 2007)

The supplier selection process overall objective is to recognize suppliers that have the biggest potential to meet company's demands consistently at justifiably cost. The selection process is about comparing different suppliers utilizing different of measures and criteria. It might not be a simple task to transform company needs to useful criteria, as needs are usually general qualitative concepts as criteria should be quantitative specific requirements. (Kahraman et al., 2003) Also in general the selection of suppliers is done based on imprecise and uncertain data that is another challenge for the selection process (Chen, Lin and Huang, 2006). What even further makes the process challenging is that the decision is usually affected by multiple conflicting factor. (Amid et al., 2009).

3.1. Supplier selection process model

Sonmez (2006) has made a step-by-step model for the five steps of the supplier selection process and describes the process in more detail (Figure 6). Supplier selection includes two different main tasks to make the initial supplier choice: 1) evaluation and assessment process and 2) evaluation and assessment aggregation. The evaluation and assessment task includes identification and decision of the selection criteria towards which the possible suppliers are evaluated and chosen. After that the evaluation metrics and scales are decided to be able to measure the supplier's appropriateness. The scales and metrics are needed to define the likely positive and negative outcomes for each criterion. (Sonmez, 2006) Depending on the particular situation, preferences, objectives and company needs the different criteria have a different importance. That is why the criteria need to be weighted based on their importance for the company. (Amid et al, 2009) Also because the environment is dynamic, weights might have to be adjusted at some point and new criteria added to be able to manage the risks in line with the latest pressure (Foerstl et al., 2010).

Figure 6. Supplier selection process steps (Sonmez, 2006)

The last step in the evaluation and assessment is to evaluate the possible suppliers against the criteria recognized at the start of the process utilizing established metrics and scales. After possible suppliers are granted the ratings or scores for the criteria, it

is important to aggregate the scores and/or ratings (quantitative or qualitative) to make a rational final supplier selection decision. (Sonmez, 2006) However Gencer and Gürpinar (2007) emphasize that it is not expected of the supplier to be totally perfect and fill all of the selection criteria company has set. For instance one supplier's cost might be the lowest but at the same time the delivery time is very bad. That is why companies have to decide what criterion is the most important for them to follow and select the suppliers based on this decision.

3.2. Supplier selection criteria

There is a very vast amount of supplier performance metrics and criteria that can be utilized for supplier selection. There are also huge amount of different decision-making models to finally choose the suppliers between the potential candidates. (Mohtasham et al., 2016) According to Kokangul and Susuz (2009) deciding on the selection techniques and criteria are the most important decisions to be made when selecting suppliers. What companies also need to decide is whether to have multiple sourcing or single sourcing as this has a high impact on supplier selection problem (Guneri et al., 2009). Overall the supplier evaluation and selection criteria have slowly become more comprehensive, systematic and diverse and are more and more based on consideration of many different aspects. The decision is not based only on single evaluation criteria such as supplier cost, price and quality. Companies consider many other factors such as environmental, service and cooperation in the selection process. Simultaneously the process is becoming a combination of qualitative and quantitative factors compared to previous supplier selection that has been originally been made as qualitative decision. Both commercial factors (price, quality, quantity, delivery time) and supply chain risks have to be considered in supplier selection. (Rao, Xiao, Goh, Zheng and Wen, 2017) According to Hallikas et al. (2004) companies should also evaluate the future when selecting suppliers; how suppliers' current knowledge, resources and orientation should be modified and maintained to be successful in the future. This is why many companies think supplier's development ability and flexibility as important criteria when selecting suppliers since markets and products are constantly changing.

3.2.1. Commercial criteria

According to Kahraman et al. (2003) criteria for the selection of suppliers can be divided into four separate groups: supplier criteria, service performance criteria,

product performance criteria and cost criteria. Supplier criteria determine if the supplier fits to company's strategy. Supplier criteria measure supplier's business such as capability and management approach, financial strength and status, quality systems, support resources and technical ability. Kokangul and Susuz (2009) add to this list the supplier's performance history and capacity but also supplier cost management. Chan and Kumar (2007) state that also the past history and performance of the supplier needs to be addressed when selecting supplier. These are for example performance history, customer base, product facility and capacity. The supplier selection that is done globally can also include ethical and environmental guidelines if they are important for the company's objectives.

Service performance criteria can be utilized to estimate the advantages generated by supplier based on company's expectations. (Kahraman et al., 2003) According to Rao et al (2017) these are also quality, quantity, and delivery time. Quality related attributes are for example product and service rejection rate, lead-time and regular quality assessment done by the supplier. Chan and Kumar (2007) state that criteria affecting service performance are also delivery schedule, R&D and technological support, ability to change and ease of communication (Chan and Kumar, 2007). Product performance criteria can be used to estimate purchased product's usability and to investigate important functional characteristics of it. The accurate criteria depend on the product. If the product is not yet developed, company has to investigate if the supplier has the knowhow to develop the service or the product. (Kahraman et al., 2003) The cost criteria include elements of cost related to product purchased such as transportation cost, purchase price and taxes (Kahraman et al., 2003). According to Chan and Kumar (2007) costs can include also product price, freight costs and tariff and custom duties. Other commercial criteria are according to Chan and Kumar (2007) for example government stability, but also the risks that each supplier bears that will be investigated next.

3.2.2. Supplier risks

Companies should strive to recognize the both the supplier specific risks and potential risks in specific countries and regions of supplier location such as shifts in political policy, currency fluctuations and other changes in the market (Kahraman et al., 2003). According to Ruhrmann et al. (2014) supplier risks can be broadly divided into two groups based on their original source, exogenous risks and endogenous risks (Figure 7). Mohtasham et al. (2016) have named these same risk categories as external and

internal risks. In addition Zsidisin (2003) has named these same risks as individual supplier failures (internal risks) and market factors (external risks). These risk classifications can also be used for supplier selection as criteria. Supply risks always have negative outcomes if they do realize such as incapability to fulfill customer needs and demand. According to Zsidisin (2003) the supply risk meaning can be different based on for example the outcome, industry and source. There is a massive amount of different supply risks and different categorizations of supply risks. There is no one unified answer to this either, which is the correct categorization to be used. It depends on the author and their opinions but also on the research context and the company's situation.

Figure 7. Supplier risks (Ruhrmann et al., 2014)

Exogenous risks are sensitive for continuous external influence so companies cannot influence or control them (Ruhrmann et al., 2014). These are for instance market and technology turbulence that arise from outside the supply chain. Some of them are really hard to predict at all beforehand. (Trkman and McCormack, 2009) Basically external risks come from interactions between the environment and the supply chain (Torres-Ruiz and Ravindran, 2018). External factors can be divided into government policies, laws and regulations and environmental factors. Also macroeconomic risks and market risks are exogenous risks. (Mohtasham et al., 2016) According to Blackhurst et al. (2008) societal risks such as political instability, civil conflicts, political events, contagious diseases, terrorist attacks, strikes and environmental disasters such as tsunamis are also external risks. Other external risks are for example economic condition and trends or geographical location (Chan and Kumar, 2007). Rao et al (2017) define economic risks as changes in the business of supplier such as price index and inflation rate changes, fluctuations in stock market or financial crisis, raw material price changes, demand changes and competitive behavior in the market. These have a direct impact on supplier cash flow and investment. Price and currency

changes also have an impact on the supply chains (Ruhrmann et al., 2014). Even though companies cannot control their external risks at all, they still should have at least risk mitigation plans and strategies in place for them (Blackurst et al., 2008). That is why it is needed to investigate the supplier location if these areas have the possibility of natural disasters and if the supplier has contingency plans and risk prevention measures in place (Rao et al., 2017).

Ruhrmann et al. (2014) state that endogenous risks or supplier-specific risks are those that are based on the performance capability of the supplier. These are communication and financial capability, confidence, quantity, bad product quality, motivation and capability. Also bad time, cost and pricing management are internal supplier risks (Hallikas et al., 2004). According to Torres-Ruiz and Ravindran (2018) internal risks arise from the relationship, cooperation and interaction between different parties in the supply chain. These risks are a result of a lack of ownership and visibility, self-inflicted chaos, JIT practices and false forecasts. Other internal risks can arise for instance from human resources and other resources, financial and IT systems and R&D (Mohtashamet al., 2016). According to Zsidisin (2003) individual supplier risks are for instance inability to handle demand and customer deliveries and technology changes and delivery problems or capacity constraints and being dependent on one supplier and not having a replacing one (sole sourcing situation). According to Rao et al (2017) also ethical risks are internal risks that are supplier's bad behavior such as cheating, fraud, leaks, distortion or unhonoured contracts and asymmetric information. Another one is management risk that is caused by poor manager quality, poor logistics and order management ability. Education level of managers is one criterion that can be used to assess this risk level. For example information risk causes easily very unsuccessful collaboration relationship with suppliers. These are for instance information asymmetry and information disclosure. The information accuracy is highly dependent on supplier's information gathering platforms and systems but also forecasting abilities and security systems in place. According to Hallikas et al. (2004) suppliers have the responsibility for confidential information. Mohtasham et al. (2016) emphasize that companies can directly and proactively strive to control internal risks and they should.

4. RISK MANAGEMENT IN SUPPLIER SELECTION

Uncertainty and risks are present in the supplier selection process. (Patra and Mondal, 2015) The capability to recognize suppliers having a higher potential of risks and disruptions is critically important for companies since supplier risks have a massive effect on the supply chain (Trkman and McCormack, 2011). Naturally when companies become more dependent on each other, companies also become more exposed to the other companies' risks. Operations that produce favorable impacts often include risks. (Hallikas et al., 2004). Also according to Rao et al. (2017) the supply chain structures are constantly becoming more and more complicated and this will lead to increased supply chain risks. When selecting suppliers companies need to take into consideration not only the commercial factors such as price and lead-time but also the risks to be able to measure suppliers comprehensively. (Rao et al., 2017) Especially when supply chains are facing risks and uncertainties, it is even more crucial to manage the risks beforehand already in the supplier selection phase (Mavi et al., 2016).

According to Patra and Mondal (2016) selecting the appropriate supplier is not an easy task. Suppliers have diverging business characteristics and each of these characteristics pose some risk behind them. These risks need to be evaluated and weighted when making the selection decision. Basically the objective is to select a supplier with minimum risks. (Patra and Mondal, 2016) Also Sanayei, Mousavi and Yazdankhah (2010) and Zeydan, Colpan and Cobanoglu (2011) state that the overall target of the selection is to minimize risks. Trkman and McCormack (2009) state that when selecting suppliers companies need to consider and estimate the supplier-connected risks and take into account the supplier market turbulence, environment, technology, supplier characteristics, but also supply chain strategy and structure. However of course these are not enough to control external supplier non-performance risks. Companies should also have some kind of risk mitigation methodology in place for these risks. (Trkman and McCormack, 2009).

According to Mensah, Merkuryev, Klavins and Manak (2017) modern supply chains are vulnerable to much bigger risks than managers of supply chains are even able to recognize anymore. The amount and types of risks that supply chains encounter are now greater than before and even the risk management has become more complex than before. If not managed appropriately, companies can lose revenues, profitability and competitive advantage to the risks and uncertainties. According to Foerstl et al. (2010) the earlier the companies start to assess their suppliers for risks, the bigger the

risk management capabilities compared to their rivals will grow to be. There is also always the first-mover advantage for the companies using established processes.

According to Mensah et al. (2017) the top three supply risks are unplanned telecommunications and IT outages, data breach and cyber attacks and adverse weather. Also transport disruptions and network were seen as major supply risk sources. If the risks realize and thus change supply chain, returning the supply chain to its original condition is a big challenge and that is why companies need to have a solid understanding of the impacts the risks have in their supply chain. According to Tummala and Schoenherr (2011) the risk categories in supply chains are disruption risks, system risks, demand risks, inventory risks, delay risks, transportation risks, manufacturing (process) breakdown risks, physical plant (capacity) risks, supply (procurement) risks, and sovereign risk. There are multiple different reasons behind the realization of each of these risk categories. The supply chain risks are split into four different categories according to Manuj and Mentzer (2008) by the influence zone these risks have on supply chain (Figure 8). These risk categories are supply risks, demand risks, security risks and operational risks. Also Yang (2011) has categorized three different supply chain risk types from the security perspective that are physical risk, financial risk and operational risk.

Figure 8. Risk types in the supply chain (Manuj and Mentzer, 2008)

4.1. Risk management process

According to Zsidisin (2003) the supplier risk management process starts with supplier network mapping. Then companies need to recognize the risks, their current location, assess the risks and manage them by constructing a collaborative supplier network risk strategy. According to Tummala and Schoenherr (2011) the risk management of supply chain has usually the following steps: risk identification, risk assessment and measurement, risk mitigation and contingency plans (risk management decisions and

implementation) and risk control and monitoring via data management systems (Tummala and Schoenherr, 2011). Every company does business at its own risk and they should manage their risks themselves. However as companies in supply chain are dependent on each other it might be useful to partially share the risk management and develop some collaborative strategies to risk management. (Hallikas et al, 2008)

4.1.1. Risk identification

According to Blackhurst, Scheibe and Johnson (2008) risk identification is a subjective but fundamental part in the process. Every company must identify the risks from the company's perspective. After risk identification their probability and impact must be assessed carefully. (Blackhurst et al., 2008) Risks must be clearly recognized and consequences comprehended so that strategies for risk mitigation can be executed. Also the understanding the interrelations, consequences and impact magnitude and variety of supply risks is important because some risk strategies can have an impact on other risks. (Tummala and Schoenherr, 2011) What needs to be noted is that companies have to determine their own risk categories based on their preferences, type of supply chain and industry and so on. There is no one unified approach to risk assessment. (Blackhurst et al., 2008)

By identifying the risks companies become conscious about the phenomena and events that cause uncertainty and risks. The main task of risk management is to identify the future risks to be able to control them with proactive approach. In a networked environment such as supplier network the risk identification phase have to take into account the dependencies on other parties within the supply chain. Delivery fluctuations, interruptions and quality failures are significant signals of production systems risks. However it is not an easy task to identify all risks and feedback loops and dependent events chains can cause this. (Hallikas et al., 2004) Event tree analysis, check sheets, supply chain mapping, failure mode, fault tree analysis, Ishikawa cause and effect analysis (CEA) and failure mode and effect analysis (FMEA) are approaches that can help in risk identification (Tummala and Schoenherr, 2011).

4.1.2. Risk assessment and measurement

Risk assessment and prioritization on the other hand are important in order to select right management actions for the risks recognized according to the company's situation

and their level of supply chain. The assessment phase can be somewhat subjective as it is based on the company's experience and also the supplier network effect has to be taken into consideration. The possible risk consequences have to be assessed from the company's viewpoint since a change or event that is harmful to one company might be positive to another company in the same supply chain. However the main target of risk assessment is to get a thorough comprehension of the environment and support management with new information. After risk assessment if can be useful to make a risk diagram of them to give an overall picture of the risks. (Hallikas et al., 2004)

Risk assessment is about the risk likelihood determination. Risk assessment can be done from objective information but if it is not obtainable subjective judgement, information and beliefs can be utilized. Some methods such as five point estimation, Monte Carlo simulation, parameter estimation, expert focus groups, probability encoding or Delphi method can help the risk assessment. (Tummala and Schoenherr, 2011) According to Blackhurst et al (2008) companies also need to weight the importance of each of the risk they are exposed to. The weights can be based on the risk probability, the relative impact of the risks, or other factors that are important for the company. After the risks are weighted the risk rating and ranking of the suppliers must be done based on their risk level. Tummala and Schoenherr (2011) state that after risk ranking and classification done based on the ranking, companies must decide acceptable risk levels. It could be good to have senior management and cross-functional teams involved in this phase and to use all available information when making this decision. (Tummala and Schoenherr, 2011)

4.1.3. Risk management decisions

In a supplier network environment the risks identified and assessed can be usually managed by having some common strategy, contract policies and best practice modes of action. Some risks can be managed by collaboration but some risks companies have to manage themselves. The basic idea on supplier network level can be to find the optimal risk management strategy level to balance and share risks between the supply chain partners. (Hallikas et al., 2004) Risk management strategies can be for instance buffer inventories, efficient relationships with suppliers and customers, information technologies, risk pooling and what if analyses, supplier contracts and multi-sourcing. (Tummala and Schoenherr, 2011) Other risk management strategies can be for instance taking risk, eliminating it, transferring it or reducing risk or having a further analysis of individual risks (Blackhurst et al, 2008)

4.1.4. Risk monitoring and control

Also risk monitoring is important as the environment around the company is constantly changing so also the risk status changes. That is why the risks have to be monitored to notice the possible increasing risk trends. Also new important risks might appear. Changes in customer needs, network, partner strategies, competitors and technology must be monitored and assessed in order to update the risk management if needed. The process of risk management should be constant and all enterprises in the supply chain have to monitor their business operations and environment, and make planning and decisions to manage risks. Risk re-assessment should be always done again when there are signals of business environment changes or when some risks have already realized. (Hallikas et al, 2004) Also Tummala and Schoenherr (2011) highlight that the risk response actions must be continuously developed to be able to control the risks and correct the risk planning based on the changes happening in the environment. Data management systems are useful tools in this.

4.2. Connecting processes of risk management and supplier selection

Supplier selection and development but also portfolio management and supplier risk management are very significant functions for efficient supplier management and selection. (Torres-Ruiz and Ravindran, 2018) According to Xiao, Chen and Li (2012) best way to integrate risk factors and other criteria to supplier selection process is to use some method that is integrated or hybrid and consists possibly of few methods and models. By doing so multiple attributes and criteria can be evaluated to make a decision, both subjective and objective risk attributes can be considered when making the decision and there is the feedback and dependence among the criteria. It is important to take also risk into consideration when selection suppliers.

According to Foerstl et al. (2010) companies should implement a structured and comprehensive supplier risk assessment, selection and development to efficiently manage the supplier portfolio. Without structured approach companies can only try to manage the supplier risks randomly which can lead to bad consequences. Risk assessment methods and technologies give companies capabilities for supplier

selection because they inhibit potentially high-risk suppliers from getting into the company's supply base in the first place.

What is even more is that risk assessment makes it possible for company to categorize its suppliers based on their capabilities and thus manage the suppliers' performance and development. By doing so it makes possible to allocate supply risk mitigation processes where they are needed most. Companies also must integrate their external responsiveness, supplier management and supply risk management to build a capability that is dynamic. This ability makes it possible to continuously assess supplier risks and monitor the suppliers. External responsiveness has a positive impact on identification, assessment and mitigation strategies of risks that in turn have a positive impact on operational performance and risk management. (Foerstl et al., 2010)

5. INTEGRATING BIG DATA ANALYTICS INTO SUPPLIER RISK REVIEW RELATED TO SUPPLIER SELECTION

Lastly in this part of the thesis, the process of selecting suppliers, risk management process and big data analytics are tied together to form a synthesis. After investigating and researching relevant topics related to supplier selection, risk management in supplier selection and big data, this thesis now strives to integrate the three subjects together. The usage of big data analytics on supplier risk management and process of selecting suppliers can have a massive effect on the overall performance and efficiency of the company. There is a huge research gap in the existing literature to studying big data analytics, supplier risk management and supplier selection together.

Today big data analytics have gained a lot of attention from enterprises (Wang et al., 2016). By collecting and analysing enormous amounts of data, big data analytics tools assist enterprises to make decisions a lot more efficiently than ever before (Militaru et al., 2015). Big data truly gives supply chains a greater data clarity, accuracy and new insights that can lead to better intelligence across the supply chain - big data can be the intense force for driving supply chains ahead (Tiwari et al., 2018). Big data gives enterprises and supply chains ability to process uncertainty and variability and provides them with advanced predictive power. Supply chain analytics derived from big data helps decision makers and managers in comprehending changing markets, recognizing and evaluating risks in supply chain, and leveraging supply chain capabilities to shape implementable and edge-cutting strategies that enhance supply chain profitability and flexibility. (Wang et al., 2016)

From risk management point of view the usage of big data analytics can lead for instance to substantial risk management cost savings due to the fact that technologies based on big data can assist risk management teams to get more precise risk intelligence in real-time. At the same time the predictive force and risk model stability is increased. (Veldhoen and De Prins, 2014) Big data can help companies to define the preventive measures to deal with risk scenarios with success (GEP, 2018). Big data analytics enables managers to be more conscious of external events possible in the future that are ever more complex in global supply chains. Big data enables global supply chains to have a proactive approach to supply chain risks which are for instance

supply failures due to natural or fabricated hazards, contextual and operational disruptions (Tiwari et al., 2018).

Big data can especially be used to enhance company's supplier selection and risk review process. Big data makes it possible to make reports on the supplier base. Big data helps in identifying business risks that are related to the company's supplies. These risks are for instance compliance risk, pricing risk, disaster risk and geographical risk to name a few. With for example data mining companies can better understand their suppliers' performance. For instance, slice-and-dice analysis that gives data based on various criteria can help companies to investigate spend information by supplier. (GEP, 2018) With big data enterprises can nonstop get up to date information on markets and suppliers and react fast to supply risks and changes. (Wang et al., 2016)

5.1. Challenges of big data analytics usage

The main challenge of using big data analytics is to analyze big data in a way that brings value. (Sivarajah et al., 2017) When companies start to use big data technologies and solutions in their operations they need to be aware of all the possible risks and challenges that are connected to the usage of big data in their business operations to be able to get value. Managers have to assess and consider, and strive to reduce the risks beforehand that may occur when using big data technologies in their business. (Raguseo, 2018) Otherwise obstacles and challenges to big data usage can have the potential to generate resistance to the big data solutions implementation if they are not considered beforehand. (Militaru et al., 2015).

There are many opinions between different authors about which are the most significant big data related challenges that will be handled more extensively below in the following sub-chapters. According to Meraglim (2017) the biggest challenge and obstacle to big data usage is data protection and security. Other challenges are lack of knowledge and access to expertise, budgetary restrictions and prioritizing other programs first. Smaller challenges are technical problems such as immature technology. According to Militaru et al. (2015) new technology investments in big data, privacy, security, protection of intellectual property and company vision are the most important reasons that might prevent the urge of enterprises to execute big data solutions in their business operations.

Sivarajah et al. (2017) categorized the big data challenges to three separate groups: data challenges, process challenges and management challenges based on the data life cycle (Figure 9). Data challenges stem from the big data characteristics: data volume, velocity, variety, variability, veracity, volatility, visualization, value, quality, discovery and dogmatism. Process challenges are a series of how questions: how to catch data, how to combine and modify it and how to find the correct analysis model and how to generate results. Management challenges include for instance security, privacy, ethical aspects and data governance. Also the lack of skills and understanding of analyzing data is a managerial challenge. Managerial problems are also data ownership, cost and operational expenditures and data and information sharing.

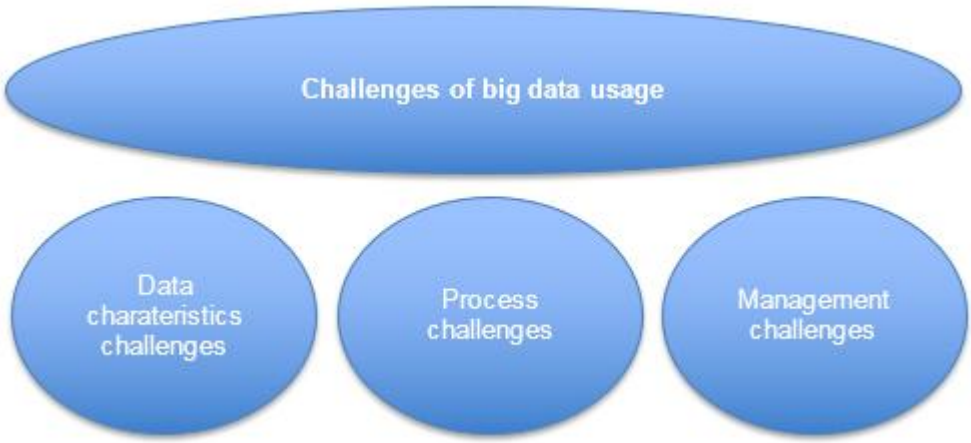

Figure 9. Challenges of big data usage (Sivarajah et al., 2017)

5.1.1. Data characteristics challenges

The complex and challenging nature of different big data characteristics cause many challenges for companies utilizing big data analytics. The volume characteristic causes homogeneity and data noise. This high volume generates algorithmic instability and high computational costs. The variety characteristic of big data demands various methodologies and techniques to handle it. All of these challenges lead to statistical biases, heterogeneity and experimental variations that make the analysis more complicated. (Wang et al., 2016) Heterogeneity and diverse data forms might prevent the analytics of big data to forecast production capacity, customer needs and demand (Militaru et al., 2015). According to Yaqoob et al. (2016) the exponential growth speed of data is a significant big data challenge because it surpasses the humans' existing capability to design suitable analysis systems and data storage systems to handle huge volumes of data efficiently. All in all the big data applications management is

these day a demanding assignment. Integrating huge and complicated big data sets is a massive challenge for companies (Sivarajah et al., 2017).

Also the value of data in decision-making might be influenced if the quality of data is bad (Tracy, 2010). Bad data quality can prevent gathering and analysis of big data (Militaru et al., 2015). The data quality can be for example its high level of processing or data relevance. Naturally there are various challenges related to big data quality as big data is such a huge and complex phenomenon. Because of the heterogeneity of data the data can be sometimes inconsistent. Challenges can be divided to engineering tasks challenges such as handling data at an unimaginable scale and semantics challenges such as discovering and meaningfully uniting information that is relevant to company needs. Sometimes it is hard to identify relevant information pieces in messy data that has also a lot of irrelevant data in it. (Emani et al., 2015)

5.1.2. Process challenges

Companies face the big data process challenges when they start analyzing and processing big data. It is hard to capture the data to interpret and present the end results gained from the analysis process because huge data sets are normally unstructured or non-relational by their nature. One of the data process challenges is data acquisition from multiple sources and warehousing it for value producing purpose. (Sivarajah et al., 2017) It is a big challenge to save huge data sets and to share data from multiple sources. Also the faster the received data information is being responded to, the greater is the benefit gained. The speed might in fact be very vital and important but also a huge challenge to overcome when it comes to big data analysis. (Salo, 2013, 71)

A great challenge is also selecting right applications and algorithms used in the big data analysis (Salo, 2013, 71). It is complex to gather valuable and usable information from big data; companies would need smart filters and algorithms for this to reduce the amount of data before warehousing it. Challenge of data cleansing and mining is related to extracting data from a huge amount of unstructured data that cannot be used as it is. Companies need a way that picks the needed data information from unstructured big data and offers it in an understandable form. Developing and maintaining this kind of method of extraction is a continuous challenge for companies. (Sivarajah et al., 2017) According to Bikakis (2018) it is already challenging to offer an overview of small datasets. Information overloading or overplotting is a general

problem also with smaller datasets but naturally the problem is even greater with big data sets.

Big data usually includes different online activities such as Facebook updates and Tweets that have different senses and meanings. This kind of vague data does not have information that can be used. Aggregating this kind of data goes beyond the capabilities of the current data integration systems. That is why useful integration of data to generate new knowledge is a key challenge in big data analytics. Data analysis and modeling after data is captured, stored, mined, cleaned and integrated is still a challenge. Old and outdated data applications for data modeling are not competent for big data analysis. Companies simply do not necessarily have good existing analysis models and methods or data processing systems. The data acquisition and warehousing can be very challenging as is data mining and cleansing let alone data aggregation and integration, analysis, modeling and data interpretation with old data processing models. (Sivarajah et al., 2017)

5.1.3. Management challenges

Often enterprises lack the digital information technologies, adequate technical skills in-house and experience to carry out big data solutions and they do not understand how to utilize analyzing techniques to run better performance in the supply chains. (Militaru et al., 2015) Also the hardware equipment needed for data analysis can be highly expensive from the management point of view even with the cloud computing methods in the market (Sivarajah et al., 2017). Companies might have the right computing technologies and systems in place but what companies need to even be able to exploit the technologies is human talent, expertise and skills (Sivarajah et al., 2017). However there is not that many of these people that can do big data analytics (Blau and Gobble, 2015). Also Bikakis (2018) emphasizes the fact that companies are facing a lot of difficulties in changing someone who is interested in big data into someone who can analyze big data. There are still a lot of people with no or little support and expertise in big data analysis. (Bikakis, 2018) The lack of supply chain integration and the shortage of big data potential vision prevent efficient cross-organizational utilization of big data and analytical tools. This kind of power (analytical skills) is very hard to produce among company employees. (Militaru et al., 2015)

Also according to Tien (2013) many managers have both a shortage of capability to see the value of big data and how to unwrap the value (Tien, 2013). However in order

for big data analysis to succeed companies' IT managers need to support existence of formal big data strategy. Enterprises have to implement and develop a right infrastructure for processing and managing big data. (Sivarajah et al., 2017). Keeping the data quality can be a hard task for enterprises in all data analysis types. That is why enterprises need good data governance systems and technologies such as updated data processing applications and data management systems that ensures clean data to handle the issue of data quality. (Yaqoob et al., 2016)

One big usage barrier of big data is the security and the legislation related to it (Salo, 2013, 24). Big data can also be a threat to privacy of users as it makes it possible to gather a lot of information about for example Facebook users and makes them traceable. How to preserve privacy is a major challenge. Privacy, security and ethics have always been one main challenge in big data management. (Emani et al., 2015) People are becoming protective of their personal data and information so security requirements are really important (Engel et al., 2016). Big data problem solving while strengthening the security is very significant problem for enterprises. Sometimes when data is produced at vast pace the recognition of the bad quality data in real-time is hard. The challenge is to ensure a real-time security protection for the data because the existing traditional security technologies are based on dataset that are static while big data is changing constantly. It becomes very challenging to design and execute novel security techniques that are able to secure the data and not cause more delay in the processing of data. This is because big data has real-time and complex processing demands. (Yaqoob et al., 2016) Even though big data is more complicated than standard data it is as vulnerable to attacks as standard data. (Sivarajah et al, 2017)

Data governance is a way to guarantee data quality and to support in attaining insights from big data in business operations. Governance is about determining where data is stored, analyzed and assessed to ensure data information quality. However significant managerial challenge related to big data governance is dividing, mapping and modeling the data after it is stored because of the complex and unstructured nature of data. Thus, efficient governance of big data is important to guarantee the analyzed and mined data quality. Another managerial challenge related to big data is information sharing. It is an enormous mission to share and integrate key data and information between organizations in supply chain. Also organizations are usually unwilling to spread their patented data from their own data warehouses because of privacy and security. The managerial challenge is to use and collect big data and respect the privacy. Also the data ownership is a managerial challenge. This is issue is specially related to social media data such as MySpace, Twitter and Facebook; does the user or

for example Facebook own this data. General opinion is that both own the data. From the ownership issue rise the problem of ensuring and controlling the accuracy of data – whose responsibility is that. Data ownership is a deep social issue. (Sivarajah et al., 2017)

5.2. Requirements of big data analytics usage

To be able to tackle the above handled challenges related to big data companies need to have certain skills, knowledge and technologies in place. Especially right personnel, skills, knowledge and technologies are important requirements companies have to be executed before they can utilize big data with success in their business. To optimize supply chain management functions utilizing big data, organizational and technological changes have to be implemented across the whole supply chain to ensure information sharing is possible between the supply chain partners (Engel et al., 2016). Because big data usage is seen hard in enterprises, supply chain partners seeking to utilize big data analytics need more than just novel skills, technologies and tools. In fact they have to think through the way they function and alter their business processes. The enterprises can benefit from implementation of big data only by innovating in processes and operations. (Militaru et al., 2015).

Enterprises are heavily dependent on strong data cleansing and data collection that are significant functions in utilizing different big data methods and techniques for analysis. That is why companies need to make substantial technological infrastructure investments to computerized systems and human resources to get data analysts on board. (Wang et al., 2016) Also according to Zhong et al. (2016) companies need to have storage technologies in their use. Salo (2013) states that data needs to be stored for the usage in the future. That is why enterprises need big data storage solutions and technologies to be able to utilize big data. Companies also need to have data transfer technology to be able to transfer data from one place to another. (Salo, 2013, 28) Although big data technology is easily accessible to anyone these technologies require application development, special skills and experimenting before they can be used for business purposes. In other words the foundation for the basic analytics technology is available but it needs to be developed further. (Salo, 2013, 95) To conclude companies basically need to have new techniques and tools but also big data enabled business models to analyze big data more easily and cheaper (Manyika et al., 2011). Some techniques companies can use are for example cloud computing and parallel computing (Akbay, 2015). Vast percentage of these tools is parts of projects called

Apache and are developed around the well-known Hadoop platform (Emani et al., 2015). According to Raguseo (2018) companies use also for example Natural Languages Processes -tool to analyzing a massive data amounts for many different objectives. Other techniques are for instance SkyTree Server, Talend Open, Storm and Splunk (Yagoob et al., 2016).

There is a forecast that there will be a massive growth in the big data skills demand before the year of 2020 (Yaqoob et al., 2016). The exponential data growth creates a huge demand for experts and specialists that can analyze the data by using mathematical models to sort out the useful data, patterns and connections in data and after that give suggestions on how to use the information to make decisions. (Blau and Gobble, 2015) Companies also need no-technical know-how and very often that is stated to be more important than technical know-how. Data analyst is a profession where the main know-how is not related to the mastery of single platform, product or technology but to the ability to comprehend business problems and formulate them into questions that big data tools can process and answer. Of course at the execution side the concrete know-how is needed but a lot more valuable is to act as a link between the problems and solutions. In data analyst the know-how of business, analytics and technology needed are rolled into one. That is why data analysts are very rare at the moment. (Salo, 2013, 75) The increasing data scientists need in Europe, US and Asia mirrors the fact that companies are at the moment wrestling with big data information (Blau and Gobble, 2015).

In the supply management and the data analysis context there are multiple enterprises that in ideal situation share their data during business transactions with supply chain partners. However enterprises should first reinforce their internal integration before integrating with customers and suppliers. (Militaru et al., 2015) Few researches made suggests that supply chain enterprises get the most gains from utilization of big data when the partners in the supply chain are symmetric in their capabilities and capacity. Of course this also depends on the types of supply chains they have – decentralized or integrated supply chain. (Bernstein, 2007) With the right kind of big data partners in supply chain it is possible to forecast precisely demand, capacity and inventory. Collaboration and knowledge sharing with big data analytics can make the decision making better in supply chains and create value. (Militaru et al., 2015) However for example data warehouses preserve huge masses of data that is sensitive and companies have to ensure that they possess a reliable security infrastructure in place that allows employees to view only information they need and are allowed to see. In

addition companies needs to have some standard privacy laws and regulations. (Sivarajah et al., 2017)

Enterprises also need the right mindset. Companies need to have the lust and courage to familiarize and examine the possibilities of novel and more intelligent analytics. In practice this requires enlightened company management and decision makers that understand the necessity and significance of technology and have the desire to try new things. (Salo, 2013, 92) To beat the obstacle of complexity, a lack of analytical skills and talent, inadequate IT abilities, a shortage of vision about the big data potential, and concern over costs the supply chain enterprises need have to better visibility, collaboration, sharing of knowledge, and optimization of inventory and logistics to gain benefits in the long term. (Militaru et al., 2015) It is important to have a vision for example that big data will be made a competitive advantage. The vision should be concrete, limited in time, understandable, realistic and inspirational. The next step from vision would be strategy but that is something that many enterprises lack when it comes to big data utilization. Companies should have as a concrete desire to use all the accessible data and see this as a possibility and not threat. The whole company should take a part in big data strategy. (Salo, 2013, 125-126)

5.3. Big data analytics methods and techniques

There are many big data analytics related methods and techniques that can be used for supplier risk review when selecting suppliers. Those will be presented in this chapter. According to Tiwari, Wee and Daryato (2018) supply chain analytics help in controlling supplier performance and supply risks. For example Fan et al. (2015) suggested a framework to notice the supply chain risks before they realize by using external and internal data. For instance social media and public news might inform exchange rate changes and disasters that can have an effect on supply chain. The framework makes supply chain risk management, monitoring, decision support and emergency planning possible in real time. Also Schlegel (2014) had an idea to use predictive analytics and big data to control risks in supply chain. Predictive analytics and big data could make it easier for supply chain managers to assess, identify, mitigate and control risks. News stations playing news around the clock, cellular communications and social media provide immediate access to information and incidents happening around the world. Also different IoT technologies such as RFID tracking and sensors can be utilized to give useful information that can be utilized to recognize and mitigate risk in the supply chain process. After these different methods

are utilized to generate a comprehensive view of the landscape of risks, the mitigation strategies can be used in a more meaningful way. (McDaniel, 2016)

Rozados and Yjahjono (2014) present some examples of practical applications of big data analytics in supply chain management. For example information from blogs and news can be processed with natural language processing or text mining by training datasets with supervised learning algorithms to identify key terms. Also this information and supplier information can be processed with for instance suitable supervised learning algorithms and expert systems modeling. Data from IoT and historical data or information of traffic, weather conditions can be processed with for example spatial regression modelling. Companies need to see big data as a strategic asset rather than as an information asset to succeed with it. The more data is used from different sources and analyzed with different methods, the more accurate the information for making the decisions is. (Rozados and Yjahjono, 2014) By using big data analytics and risk management together as tools for the supplier selection, the process could be drastically enhanced. This chapter presents different possible supply risk management methods that have big data analytics characteristics for companies to see where they could begin to incorporate BDA across their supply chains allowing them to solve complicated problems such as supplier selection.

5.3.1. Predictive, descriptive and prescriptive analytics

Advanced big data analytics methods are predictive, descriptive and prescriptive analytics. According to Sivarajah et al. (2017) these kind of analytical methods support organizations in their performance and decision-making by enhancing them. Everything becomes more quantifiable and translucent while recognizing potential concerns, opportunities and inconsistencies. (Sivarajah et al., 2017) Both prescriptive and predictive analytics have an important role in assisting enterprises in making efficient decisions when it comes to the enterprise's strategic direction (Demirkan and Delen, 2013). They can be used to handle different challenges such as modifications in enterprise culture, configuration of supply chain, sourcing decisions, and development and design of services and products. These all are strategically important decisions. Enterprises also get deeper understanding of their business dynamics and thus can optimize their day-to-day operations. (Wang et al, 2016). Consequently, it is clear that predictive, descriptive and prescriptive analytics makes the risk management a whole lot easier for enterprises because enterprises can get insights on what will possibly

happen in the future. Companies can also suit these methods to fit into supplier selection process.

Predictive analytics is traditionally used to lessen the uncertainty of decisions and to forecast future trends and happenings. Predictive analytics betters the risk prevention and proactive decision-making. To be able to utilize predictive analytics companies must trust on predictive analytics algorithms (machine learning, neural networks and regression analysis) and a predictive analytics platform that give optimal decisions for companies. Programming and advanced mathematical algorithms are used together with computing power to uncover where the business operations are going. (Wang, Kung and Byrd, 2016) Predictive analytics uses also web/data/text mining (Demirkan and Delen, 2013). Predictive analytics investigates data patterns and relationships by utilizing simulation, statistical modeling and programming. It tries to also discover the reason behind possible happenings and uncover the information not available to companies. (Tiwati et al., 2018) Predictive modeling is basically forecasting the future happenings based on semi-supervised, supervised and unsupervised models of learning that are used to analyze historical and current data. There are two types of predictive analytics: machine learning techniques and regression techniques. Some techniques such as moving averages try to recognize historical patterns and apply them to the future. Others such as linear regression try to find interdependencies between different variables and utilize them to make predictions. (Sivarajah et al., 2017) However the longer the time period and the more there are different variables – the more imprecise the forecasts become (Salo, 2013, 28).

The descriptive analysis targets to recognize the opportunities and problems in the current functions and processes of the companies (Wang et al., 2016). Descriptive analytics strives to find out what has happened, what is happening and why it is happening by providing reports that give historical information on company's operations. (Tiwani et al., 2018) Descriptive analytics uses information and data available to qualify the present business situation so that patterns, exceptions and developments are recognized with standard reports and alerts, dashboards, data visualization and scorecards. It is the most simple and traditional big data analytics technique as it tries to model past behavior and what has already happened from huge amounts of historical data. It strives to describe and summarize patterns of knowledge with simple statistical techniques such as median, mean, mode, variance, standard deviation and frequency measurement. However most of the big data analytics is descriptive and exploratory analytics and the usage of descriptive statistical methods such as data mining makes it possible to find unidentified correlations and patterns that

can help for instance decision-making. (Sivarajah et al., 2017) Descriptive analytics happens whenever required or at standardized periods using technologies such as drill down or online analytical processing tools such as OLAP system and visualization. (Wang et al., 2016). These tools are supported by reporting technology and real-time information (Tiwari et al., 2018).

Prescriptive analytics tries to answer to question what should happen and how to affect it by making up different decisions based on predictive and descriptive analytics by utilizing multi-criteria decision-making, simulation, randomized testing and mathematical optimization techniques. However this is quite complicated and many companies do not use it in their operations on daily basis. (Tiwari et al., 2018) Prescriptive analysis uses also mathematical algorithms and data to assess and help in making various decisions that contain requirements and goals that are portrayed by complex nature and high volume, with the target to better the business performance. (Wang et al., 2016) Prescriptive analytics tries to solve the relationship of cause and effect from analytics results and policies of optimization of business process to assess how business betters their service operations while reducing the costs. Analysis helps business analysts to assess the impacts of their actions regarding business requirements, constraints and objectives and thus determine the actions that should be done in the business operations. What if simulations can help in determining the needed business actions by using feedback gained from predictive analytics. (Sivarajah et al., 2017)

5.3.2. Statistical analysis and optimization

Statistical analysis is composed of two analyses: descriptive and inferential statistics. Descriptive statistics utilizes data to summarize or describe features of a population or a sample by using tables, graphs or numerical calculations. Inferential statistics utilizes data to reason properties and to make predictions about a population. (Tiwari et al., 2018) Because of this companies can utilize both qualitative and quantitative techniques. Qualitative techniques are based on consumers' or experts' judgment that are subjective. This is appropriate approach when there is no past data available. Quantitative techniques can be utilized to do predictions from past data. Both techniques are used for intermediate- and short-range decision-making. (Wang et al., 2016)

Statistical analysis is needed by companies when they tackle for example with uncertainty in risk analysis of supply chain. Statistical multivariate methods are helpful for monitoring supply chain network to mitigate risks. Big data characteristics; volume, velocity, variety and veracity; demand easy to adapt and robust statistical analysis techniques such as data-parallel statistical algorithms to perform refined statistical analysis of big data. These algorithms strive to compare and model big data's distributions and densities with particular techniques such as Mann-Whitney U Testing among many others. (Tiwari et al., 2018) Two extensively utilized quantitative methods in supply chain analytics are time series analysis and regression analysis. Time series analysis processes data to pick up meaningful statistics and patterns. Time series forecasting predicts what will happen in the future based on historical data. Regression analysis makes it easier to comprehend causality effects and relationships between variables. (Wang et al., 2016)

Optimization method is a strong tool for data analytics in supply chain. The method is capable to get insights from a complicated dynamic system with multiple constraints and factors with vast data volumes. Optimization method is also able to reveal new data and analyze various objectives at the same time. The technique could help to leverage planning accuracy of supply chains. (Tiwari et al., 2018) Quantifiable problems are solved with using optimization methods and they are very effective because they exhibit parallelism. Optimization strategies must be used in a real-time environment so they are able to handle big data applications. (Yaqoob et al., 2016) The utilization of optimization methods in supply chain analytics assists in enhancing the punctuality of supply chain planning and forecasting the demand. Optimization can also get insights from data that let decision-makers to do suitable decisions. At the same time optimization helps in handling and analyzing the supply chain performance measures such as demand fulfillment and cost reduction. However even though optimization techniques deliver optimization but are time-consuming and have high complexity. (Wang et al., 2016) Because of optimization's flexibility it is also possible to find new connections in data and change them into information that can open more business value from data (Balaraj, 2013).

5.3.3. Modeling and simulation

Modeling and simulation have a significant part in big data analytics in the big data application development. They help the developer of the application to run "what-if" analysis under various system complexity and configuration situations. In the analytic

process, simulation makes possible the predictive, prescriptive and diagnostic analytics for decision maker. (Tiwari et al., 2018) What if analysis indicates how different scenarios and situations change the supplier risks. Companies can also adjust their what if scenarios if they want for example to make the overall risk measures more sensitive to quality changes. (Blackhurst et al., 2008).

Risk models can help in gaining a more precise view of risks and in designing better business continuity strategies (Michel and Parasuraman, 2017). A risk model is a mathematical representation of a system and it usually includes probability distributions. Risk models can help in understanding vast strategic objectives or supply chain threats, evaluate geopolitical risks of new markets or answer very specific questions companies might have. After risk models are developed, the models can be utilized to evaluate how system behaves under normal conditions but also under what if scenarios. This can help companies to determine their risk tolerance level and to evaluate how to be able to withstand different impacts. Companies using risk models have to understand how the models fit into the bigger picture of how it collects and utilizes information of risks to make decisions. (Curtis, 2016)

Simulation is the process of determining how risk model behaves under certain assumptions or conditions to avoid surprises. The simulation results can help to make decisions or they give insights into the system or processes investigated. Model can be strengthened or adjusted based on the simulation results or as the underlying assumptions or conditions change. (Curtis, 2016) Scenario simulations consider massive data amounts that allow effective realization of risk concentrations and quicker reactions to developments in market. Especially Monte Carlo simulation is flexible and effective tool for this. An enhanced accuracy is gained by a larger simulation amount the model can project. (Corea, 2016) Together modeling and simulation make decision making easier by reducing irrelevant information thus making the multidimensional problem more manageable. However companies should always remember that there is some uncertainty in modeling and simulation and the results should not be trusted blindly. (Curtis, 2016) Figure 10 demonstrates the process of catastrophe risk modeling (Michel and Parasuraman, 2017).

Figure 10. General framework of catastrophe risk model (Michel and Parasuraman, 2017)

The exposure data in catastrophe risk modeling framework could be for instance supply chain data that can contain geospatial and physical details for both enterprise's and suppliers' production facilities. The exposure data can also include historical information of supply chain disruptions at each of the supplier locations with the claims data. In addition exposure data could contain supplier tier information, mitigation measures and product codes used by each supplier (reserves, resilience and redundancy). (Michel and Parasuraman, 2017) Collecting the right data is one of the biggest challenges of risk modeling and besides relevant historical data also expert opinions are needed to understand the potential severity of each risk and its probability of occurring otherwise the decision-makers might not be comfortable enough with the risk models to take them into decision-making. (Curtis, 2016).

In the hazard phase the model calculates for every supplier location the hazard intensity from the exposure data such as hurricane wind speed, earthquake shaking and flood depth, for both stochastic (future) events and deterministic (historical) events. Then in the next phase (engineering phase) the damage functions utilize policy conditions, post-event surveys, claims data and engineering analysis to relate intensity of hazard to physical damage realized. These are estimated in every supplier location to evaluate the supplier disruption likelihood. The disruption at one supplier location also affects all the downstream suppliers with the utilization of supply chain product flow network information. One option for companies to avoid risks is the usage of the hazard data in the process of supplier selection to remove high-risk regions' suppliers or to put in place measures to mitigate the risks or protect themselves against possible problems. (Michel and Parasuraman, 2017)

Financial phase calculates business interruption losses based on stochastic and historical disruption events in the supply chain, direct business interruption losses and

individual suppliers' physical damages. The estimates of supply chain losses can be outputted as Business Interruption (BI) in days or as Product Value at Risk (PVaR) that is the lost revenue from shortfalls of production. Risky suppliers can be identified with the usage of loss estimates. Identifying critical suppliers can help companies in planning for contingencies such as getting suitable insurance cover or build up risk mitigation measures for their suppliers. (Michel and Parasuraman, 2017)

5.3.4. Visualization

Data visualization is the displaying of big data in a graphical or pictorial form to recognize unexpected phenomena from the massive data amount and perform multiple complicated operations. Visualization tools have a big importance in big data analysis era as they give companies an intuitive way to interactively analyze and examine data. This makes it possible for companies to recognize meaningful patterns, causalities and infer correlations that support decision-making operations. Companies use visualization to monitor and research markets, find opportunities in emerging markets and communities, optimize operations, develop business and analysis, perform predictions and identify trends. (Bikakis, 2018) The big data and the dynamic data visualization tools together are a good method to use data analytics to address risks (Curtis, 2016). Visualization can be used to generate diagrams and tables to make big data more easily understandable (Yaqoob et al., 2016). Also animations and images are visualization techniques used to better understand, improve and communicate the big data analysis results (Manyika et al., 2011). To be able to make good decisions and solve problems enterprises need to have a clear visualization of the big data content. Despite all the benefits there are still problems in big data visualization. The visualization is still really static and data is not stored in a relational way. Static nature of visualization is preventing updates in real-time. (Emani et al., 2015)

Modern visualization tools that can be used for big data should efficiently and effectively deal with these facts; real-time fast interaction and analysis, on-the-fly processing of raw not preprocessed data, visual scalability and user assistance and encouraging and tools personalization and customization based on customer needs. Visualization tools also need to generate effective and efficient summarization and abstraction mechanisms. Many of the visualization tools are approximation techniques that are usually defined in a hierarchical manner that makes it possible to examine data in various detail levels (hierarchical aggregation). Hierarchical data visualization tools are extremely good for massive graphs visualization because data is recursively

broken down into smaller sub-graphs. In this way the visualization offers an overview of data but also makes it possible to find specific and detailed information from the data. These kind of hierarchical visualization tools can also handle the overloading problem as they use approximation techniques. (Bikakis, 2018)

There are various techniques for big data visualization that are for instance Spatial Information Flow, Clustergram, Tag Cloud, and History Flow. According to Emani et al. (2015) Spatial Information Flow is a visualization technique. It is mainly a lighting graph where edges connect sites located on a map. Schonlau (2002) determine Clustegram as a visualization technology that is utilized for cluster analysis that displays how individual dataset members are assigned to different clusters as the number of clusters grow. The number of clusters is significant in every clustering process and it has the advantage to more easily detect how the number affects results of portioning. (Schonlau, 2002) Tag Cloud is a technique for linking and visualizing concepts of specific website or domain. The concepts are written by utilizing text properties such as weight, color and font size. (Emani et al., 2015) History Flow is a visualization technique to demonstrate the document evolution effectively with respect to the contributions of its various authors. The vertical axis of history flow presents the names of the authors and the horizontal axis the time. Every author has an individual color code and the vertical length of the bar show the amount of text written by authors. (Víegas, Wattenberg and Dave, 2004)

6. RESEARCH METHODOLOGY AND DATA

The empirical part of this thesis was made in a qualitative way and further as a case study by interviewing employees from two separate case companies. As neither of the companies utilize big data in their operations widely but rather have only investigated it a bit and know the term, the companies are selected on the assumption that they could theoretically benefit from big data usage in their supply chain management and supplier selection. The objective of the empirical part of this thesis is to deepen the understanding of the research topic and in a sense mirror the previously presented theoretical part from a practical company point of view. The aim is to further investigate how companies and especially their supply chain management departments see big data and its possibilities and challenges related to integrating big data analytics to supplier selection and supplier risk management processes. This empirical part also aims to describe how big data analytics and risk management process can be utilized together in the supplier selection to make the whole process more efficient.

Next the research methodology and data utilized in this research are explained more deeply to provide an introduction for the empirical analysis and results. Overall the goal of this research chapter is to clarify the chosen research methodology and describe how the research data was collected and analyzed to find answers to the research questions that were investigated in the theory part of this thesis. Finally, the validity and reliability of the research is evaluated, and a short case company description is given before the actual empirical findings and analysis chapter.

6.1. Research method

Because of the design of the research problem and questions, the research method for the empirical part of this thesis is a qualitative research. Qualitative research is suitable for the topic of this thesis in sense that it offers a way to withdraw from theoretical and conceptual customs that guide the mainstream researches. Especially qualitative case study research is suitable for research situations in which the subject nature is in a need of deeper understanding. (Koskinen et al., 2005, 23-25) Big data is something that is not yet investigated that much from the business world point of view, so the topic is in a need for a deeper understanding which can be reached with qualitative research. Further case study approach is suitable for the topic and research question of this thesis as case studies are used to answer research questions such as “how”

(Robson, 1993, 44) and the main research question of this thesis is *How big data can be exploited in a risk review related to supplier selection to make the process more effective?*

The case study is suitable for the topic of this thesis because by using case study it is possible to obtain understanding of complex phenomena and to gain specificity to the thesis subject. Also with case study it is possible to gain a comprehensive picture of investigated case companies. (Koskinen et al, 2005, 156) The case studies are done as semi-structured interviews alias theme interviews by interviewing people that are in touch with big data utilization or work within supply chain management in the selected case companies. In theme interview the interview is usually implemented with questions that are made beforehand by the interviewer. The interviewee can answer freely in their own words to the questions. Theme interview is used in this thesis because when conducted carefully it is very efficient way of making qualitative research because the interview can be guided without controlling it entirely. (Koskinen et al., 2005, 104-5) However the interviews made via e-mail are naturally structured interviews as they are just sent via e-mail so there is no space for naturally flowing discussion done with those interviewees.

6.2. Research data and data collection

The theory part of this thesis is constructed of scientific articles and other literature such as books and previously made researches on the topic. In addition some Internet news and articles are used that are related to the research topic. The topics of the sources are connected to big data, big data analytics, supply chain management, supply risk management and supplier selection process and supplier risk review related to the supplier selection process. In this way it has been made possible to construct a theory part for the thesis that is meant to build an extensive basis for the empirical part. The data utilized in the empirical part is from the interviews with the representatives from the two case companies investigated in this thesis. Interviews were chosen as the method to collect data as this research is aiming to get a comprehensive and thorough understanding of the studied phenomenon. The chosen interview questions are based on the theoretical part and the framework of the research so the empirical part handles the same topics as the theoretical part but from a more concrete company point of view.

The interviewees were chosen based on a few criteria. Because of the research topic most of the interviewees were from the case companies' supply chain management department. However, from the company Beta three of the interviewees were from digitalization and data management departments to be able to give their insights and more precise understanding of big data to be able to investigate further the case companies' readiness to start utilizing big data. This is because big data is not used in either of the case companies so naturally the supply chain department employees do not have that wide understanding about big data and its possible opportunities. These interviews were valuable as big data is somewhat complex phenomenon and not that many companies and their employees totally understand it, so these interviewees were able to give more understanding on the big data concept in the case companies.

In total five interviews were made in August and September 2018 in the company Beta. Before the interviews in case company Beta three interviews were made in company Alpha in June 2018. The interviewees had alternative ways to answer to the research questions. Most of the interviews were made as face-to-face interviews but few of them were made via e-mail because of busy schedules and time differences. The three interviews made in case company Alpha were conducted at the same time face-to-face with all the three interviewees. The interviewees' positions, case company they are working in, interview methods, interview questions and interview language are listed in the Table 1. All the interviewees received the research questions in English but most of the interviews were conducted in Finnish.

The interview questions (appendix 1) were selected based on the research topic, the research questions, the thesis framework and overall the theoretical part of the research. The questions were categorized under few themes according to the theoretical part's wider topic entities. Interview questions were a bit different to people working in supply management department and to people working in digitalization and data management department. For the interviewees not working within case companies' supply management department, only interview questions about the big data and integrating big data to risk review related to supplier selection were asked as they did not have anything to say to the supplier selection process or risk management in supplier selection. The face-to-face interviews were made with a semi-structured interview method. However the interviews made via-email are naturally made as a structured interview as the interviewees just answer to a set of questions given to them via e-mail without any naturally flowing conversation guiding the interview as the research questions are set as they are and there is no discussion made in the interview.

Table 1. Interviewees of the case companies

Interviewee position	**Case company**	**Interview method**	**Interview questions and language**
Global Procurement Manager	Alpha	Face-to-face	Template 1, Finnish
Sourcing Manager	Alpha	Face-to-face	Template 1, Finnish
Sourcing Manager	Alpha	Face-to-face	Template 1, Finnish
Director of Product Sourcing and Supply	Beta	Email	Template 1, English
Purchasing Manager in Project & Services (in USA market)	Beta	Email	Template 1, English
Project Manager in Digital Solutions Platform and Operational Excellence	Beta	Face-to-face	Template 2, Finnish
Manager in Smelting Automation Digitalization	Beta	Face-to-face	Template 2, Finnish
MIS and MES Data Management Specialist	Beta	Face-to-face	Template 2, Finnish

6.3. Case companies

Company Alpha is one of the leading global suppliers of services, equipment and systems in its industry. The company operates in more than 40 different countries worldwide. The company was established in 1850s in Austria. The company Alpha has about 1 200 employees in Finland and even more worldwide in its subsidiaries. Especially ensuring the quality of their services and products is strategically highly important for the company but also improving the customer's production and operation efficiency is important for the company.

Company Beta is a Finnish company and as company Alpha, one of the leading global companies in its industry it operates on. The company is established in the early 2000s. It operates in more than 30 countries around the world and has about 1 000 employees located in its Finland offices. Especially sustainability and quality are very important values for the company. As company Alpha's also company Beta's early roots are located in the central Europe and not Finland.

6.4. Data analysis

Content analysis is the data analysis style chosen to use in this study as it is a usual choice for analyzing qualitative case studies (Riessman, 2008, 194). The theoretical part of the existing theory in a way guides the analysis of the empirical data findings. The aim is to compare the empirical findings to existing theory and that way answer the research questions. The main idea of the analysis is to find the most important and relevant empirical findings from the interviews to discover deeper meanings from the answers rather than only presenting word to word what the interviewees said. The wider themes in the empirical analysis are the following (also investigated in the theory part of this thesis as wider categories or themes):

- Supplier selection process
- Risk management in supplier selection
- Big data
- Integrating big data to supplier selection

None of the interviewees wanted their voice to be recorded in the face-to-face interviews, so the notes from the face-to-face interviews are made with writing notes to a Word file as the interviewees answer to the interview questions during the semi-structured interview situation. This allows reviewing the interview answers as much as needed when analyzing them afterwards for the empirical findings and analysis. The interviews that were made in Finnish were translated into English for the empirical findings chapter. The analysis of the research data started with reviewing and reading the interview answers to find similarities and differences in the interview answers and to recognize the same themes handled in the theory part. The most significant findings are presented and analyzed in the next chapter of empirical findings and analysis. The empirical findings chapter describes the most important findings, mirrors them to the theoretical part of this thesis and at the same time themes and organizes the findings under different topics that follow the theoretical part and interview questions.

7. EMPIRICAL FINDINGS AND ANALYSIS

This chapter presents and analyzes the empirical findings and conclusively the results from the interviews conducted in the two case companies. The chapter is arranged to be corresponding with the theoretical part presented earlier in this thesis; the empirical findings are presented under wider topic entities that are the same as the themes of the theoretical part of this thesis. All of the themes handled in the empirical part follow also the research questions which guide this study. First theme analyzed in the empirical part of this thesis is the supplier selection process. Following that, the other larger themes analyzed are the risk management in the supplier selection, big data and finally the integrating big data to supplier selection process -theme. The processes of both case companies' supplier selection and supplier risk management are investigated to seek whether the case companies would benefit and have a need and space in their processes for big data analytics utilization in their supplier selection practices to make it more accurate. Also it is investigated how case companies see big data and big data analytics, and its opportunities and challenges. In addition case companies' general readiness stage for big data utilization is investigated and what changes they think they would need to make in their organizations to be able to start big data utilization in their supplier selection.

7.1. Supplier selection process in case companies

For company Alpha supplier selection is very important according to the interviewees. All new suppliers are investigated thoroughly before accepting them to company's supplier base. However the interviewees from company Alpha state that it is hard to forecast possible problems with their old suppliers. Company Alpha interviewees state that they do not utilize any specific named theoretical model for selecting suppliers, but still they have rather strict process steps they follow for supplier selection. Also for company Beta right supplier selection is very crucial process. According to company Beta's purchasing manager in product and services in USA the importance of supplier selection is depending on the item categories. For high engineering and specialized products, the supplier selection is very important as the supplier selection highly affects the performance of these products. Also for high value and volume items the supplier selection is especially important. As the supplier selection process and its success seem to be very important for both case companies, it could be very beneficial for company Alpha and Beta to investigate and estimate whether they could gain

something out of big data analytics utilization in their supplier selection and risk management process.

7.1.1. Supplier selection process steps in case companies

Company Alpha constantly screens their markets for possible suppliers according to the interviewees. Sometimes the supplier selection process starts simply from a need to find a new supplier to some market area of their business operations. The supplier selection process in company Alpha might also start when their subsidiaries in different countries want to expand to new countries. Company Alpha interviewees feel that sometimes the selection process is easy but sometimes it is harder to find possible suppliers. In summary company Alpha first selects a few possible candidates from a pool of suppliers, makes a preliminary summary of them and then investigates and audits the supplier candidates in more detail. If the result is a positive one they accept the suppliers in their supplier base. They feel that all in all company Alpha has straight forward and clear supplier selection process steps. Company Alpha also hosts so called “supplier fairs” into which they invite many potential suppliers and have conversations with them, audits them and possibly makes some new connections with them. Sometimes they arrange a bidding process to select suppliers to specific projects and in some cases they might have some contract agreements with the suppliers. Company Alpha also has an SRM system in place in which they have as much as approximately 40 000 suppliers.

As in company Alpha also in company Beta they do not have any particular method for supplier selection according to the director of product sourcing and supply. The supplier selection is rather done based on their specific business needs. According to him the supplier selection process steps in company Beta are roughly the following: defining business needs and profiling supplies accordingly, screening of potential suppliers, finding most potential suppliers by using for example RFQ (request for quotation) and clarifying interest for cooperation, having meetings with suppliers and making possible audits. The last step is naturally the final selection of suppliers into the company Beta’s supplier base. According to purchasing manager in product and services in USA company Beta just uses simple bid evaluations (comparison sheets) to select suppliers. They just select suppliers that are suitable technically and commercially and are able to deliver on time with the suitable price.

As both case companies have quite open process for supplier selection, it could be easy to start utilizing big data analytics in their supplier selection process as the companies are not too strict about the process steps. Big data utilization could be especially helpful to company Alpha and why not also to company Beta in situations when the supplier selection is more complex and harder and when they are having troubles finding any suitable suppliers with their traditional methods. For example all of the interviewees in company Alpha stated that sometimes the supplier selection is really hard. Especially company Alpha could also utilize big data analytics to narrow down their quite massive supplier base by recognizing and removing with the help of big data analytics badly performing suppliers. As company Alpha has around 40 000 suppliers traditional methods probably are not very efficient to control this many suppliers. Big data analytics could make the supplier management much easier for company Alpha because with big data it is possible to handle and process massive data amounts and around 40 000 suppliers truly create a huge amount of data.

"Who does the final decision about supplier selection? (Who has the responsibility?)"

Company Alpha does a lot of project work in their business. That is why usually the people and purchasers involved in any specific project select the suppliers suitable to that specific project. However they can only select suppliers that are acceptably audited to the company Alpha's supplier base. In company Beta the supplier selection responsibility is depending on the situation at hand according to the director of product sourcing and supply. Suppliers for company Beta's proprietary equipment are selected by specific product line employees who know a lot about the equipment and know what need to be demanded from the suppliers. However quite often also the people working in supply can make the supplier selection decision. According to purchasing manager in product and services in USA the final decision about suppliers in company Beta is made by team-specific engineering project manager but for higher value items the key management might also do the decision or at least take a part in it. With the help of big data analytics the people responsible for supplier selection could make their job and decision-making a lot easier. However of course if the case companies would start utilizing big data analytics some of the supplier selection responsibility would shift away from them to the people working with big data analytics who would have the responsibility to ensure the information derived from suppliers with the help of big data analytics is accurate.

7.1.2. Supplier selection criteria in case companies

The most important supplier selection criteria company Alpha has is the supplier's technical competence. Suppliers must be able to manufacture the company's products at the right quality level. Also code of conduct is important for company Alpha as a supplier selection criteria. They also avoid suppliers that have a collaboration relationship with company Alpha's rival company. Basically the suppliers must have good quality products and be suitable in size and price. Company Beta has some same supplier selection criteria as company Alpha. The most important supplier selection criteria according to director of product sourcing and supply is supplier's technical suitability including quality, competitiveness (price) and supplier's interest to work closely with company Beta. These are the "must have" criteria for supplier selection. The other important selection criteria are size of the company (not too big or small) and location (depending on the need). According to purchasing manager in product and services in USA company Beta's most important supplier selection criteria depend on the situation and the products needed to source. In some situations the delivery time is the most important and in some situations the price or technical parameters are most important criteria for supplier selection.

"How do you get or gather information on your suppliers to make the selection decision?"

When selecting suppliers company Alpha gathers information from possible suppliers from the Internet and from the supplier fairs they arrange regularly to find new suppliers. Also when company Alpha's employees travel the world on their business trips to their customer sites and subsidiaries company Alpha has in different countries, they might also get tips about possible suppliers. In addition also their business partners give them tips about good suppliers. All in all company Alpha gets information on suppliers from many different sources, they do not use any specific one information source when gathering information and selecting suppliers. On the other hand in company Beta according to the director of product sourcing and supply they gather information on suppliers directly from the suppliers or from public sources if they are available and from their colleagues. However the purchasing manager in product and services in USA in company Beta has a quite different view about the supplier information gathering. He feels that how company Beta selects suppliers is mostly only based on the supplier quotations and they do not necessarily gather any extra information on suppliers.

Both companies could benefit from utilizing big data in their supplier information gathering. With big data analytics it is possible to combine different information sources into one as both case companies utilize many different information sources when they seek for information about their suppliers and possible suppliers. By having all of the available information on suppliers in one place the supplier selection process could be a lot more straight forward and efficient and the companies could be able to get all the important information from massive amounts of data. Of course all information case companies use for supplier selection cannot be handled with big data such as the tips from companies' partners, customers and colleagues in subsidiaries but the information for instance from Internet and other public information sources can be handled with big data analytics. If the case companies have some agreements with their suppliers about sharing information to one another big data analytics can also be used to analyze this information.

"What kind of challenges do you encounter with the supplier selection process and supplier information gathering?"

There are some challenges related to supplier selection that the company Alpha interviewees recognize. In their opinion it is hard to get information from medium sized companies and also situations in companies this size change a lot faster than in bigger companies. Also from many companies there is not much public information available for example on newspapers or Internet. The smaller the company is the harder it is to get information on them. A big challenge is for instance to get information about suppliers' financial situation and often the information available is not that relevant or even accurate. According to company Beta's director of product supply and sourcing the challenge related to supplier selection is that sometimes it is hard to recognize all the stakeholders that should be participating in the supplier selection decision making. Company Beta's purchasing manager in product and services in USA recognizes that the challenge related to supplier selection is that the different people have different choices and opinions when selecting the suppliers also different situations can cause challenges to the process of selecting suppliers. The big data analytics could also help with some these challenges mentioned by the case company interviewees but of course if the information is simply not available even big data analytics cannot do much.

7.2. Risk management in supplier selection in case companies

Company Alpha

Company Alpha interviewees state that they do have some risk management process steps in place but they are not that specific. If they notice difficulties and problems with one supplier, they put that supplier into special observation category. Company Alpha also follows constantly their suppliers' deviation notifications, complaints and use many different systems and tools they have in place for supplier risk investigation. They also use external risk method company that offers data information on the suppliers around the world such as what risks are related to suppliers' environment. The external risk company also screens Internet with different algorithms. Still the information might be a bit simple such as increased exchange rates. The tool this external risk company uses also estimates widely all the possible problems for example a delivery they send can encounter during a 2 months delivery to customer such as storms. However according to the company Alpha interviewees the most up-to-date information they get and rely on when it comes to risk management is the information they hear from colleagues and suppliers' stakeholders. What the external risk company does for company Alpha they could also do it themselves with the help of big data analytics and make it even better.

Company Alpha interviewees feel that there is no one supplier risk that is greater than any other risks as anything can go wrong. The most common risks they have encountered are supplier quality risks and "unexpected" capacity risks which means the supplier agrees to sell more than they are able to manufacture within the agreed time and thus cannot deliver on time. Also finance risks and possible bankrupts are something company Alpha has encountered almost every year with its suppliers. This means they have had to sponsor their suppliers in the middle of the project more and earlier than planned because the suppliers simply run out of money. There are also some country specific risks such as different political risks that affect company Alpha but according to interviewees political risks are not present in that many countries their suppliers operate. However for instance in China if the state makes an order to a supplier from which company Alpha has also ordered something, the state order goes past the company Alpha order and in this ways delays the company Alpha's order. Political risks for example in USA due to their new president Donald Trump also create cost pressures and overall a lot of uncertainty to the markets. Delivery and logistics risks are also something that company Alpha has faced. These risks can happen for example when there is a storm at the sea or people make mistakes when lifting the products in the warehouses.

Company Beta

Company Beta's director of product sourcing and supply states that they manage and assess the supplier risks already when they approve the suppliers to their supplier base. He also states that company Beta has a separate project specific risk management process in place which covers the supply and suppliers. This means risk management is not done on a company level. Also purchasing manager in product and services in USA in company Beta states that their suppliers are always evaluated and screened for technical, financial and EHS capabilities before approving them to the supplier base. He also mentions that legal and financial risks are covered by terms and conditions and other financial tools such as BG and LC etc., transportation risks with transportation insurance and delivery risks they are trying to control with expediting suppliers as much as possible. According to him the project managers also do risk assessment and measurement periodically and try to provide mitigation measures to reduce risks. Company Beta could take big data analytics for instance as one part to their project specific risk management process to make it more accurate.

According to company Beta's director of product sourcing and supply quite typical risks for company Beta are related to supplier capacity constraints and most severe risks are quality risks as they can be quite expensive for company Beta if the defects are found by the end customer. Purchasing manager in product and services in USA in company Beta feels that the greatest risk they have is that they are unaware of the end market. His department for example is in USA and they are supplying to UK market, so their vendors do not always fully fathom the challenges of it. The risks company Beta has faced before according to him are especially quality issues, poor delivery and there has been plenty of back charges to them.

"Who is responsible for risk management in your company?"

In company Alpha the responsibility of supplier risk management is almost always project specific. In every project company Alpha has they assign a certain person to look after risk management. This person has the responsibility to investigate everything possible within supply chain and suppliers. They also follow up on monthly basis their suppliers in company Alpha. However the interviewees state that they do not have any specific systematic and comprehensive way for doing risk management in company Alpha and they do not do it at the whole company level. Risk management is seen in company Alpha rather as a category responsibility. Employees think only their own

responsibility areas and take care of possible risks those areas can encounter. The purchasing manager in product and services in USA in company Beta states that all of the project members are responsible for the risk management but the project manager has the ultimate responsibility.

"Do you have any challenges and problems related to risk management?"

Company Alpha interviewees feel that there are some challenges and problems related to the risk management process. According to them it is very hard to forecast and predict the risks as the information available is usually not up-to-date. Also the risks they encounter are usually cyclical. When the business is in the boom, the suppliers are usually too optimistic and sell more they are even able to manufacture as they want to sell as much as possible. On the other hand, in the recession suppliers usually make bankruptcies when they run out of resources. Companies cannot either control everything such as weather and climate. Purchasing manager in product and services in USA in company Beta notes that appropriate training for team members would help company Beta significantly to perform their risk management better. However big data analytics could enhance the risk management process even more. Big data analytics systems could offer some warning signs for both case companies to tackle the problem of surprising and cyclical risks. Also better risk forecasting could be possible with big data analytics as they could teach the systems to recognize the early warning signs of possible risks. What is more with big data analytics it is also possible to achieve the situation where the data utilized is actually up-to-date which itself would make the risk management process all in all more accurate.

7.3. Big data in case companies

Now the empirical part of this thesis will continue to handle the wider topic entities of big data and integrating big data analytics into supplier selection from a more realistic point of view. In the previous parts of the empirical chapter the main focus was on the case companies' supplier selection and risk management processes. Both processes and case companies' possibilities to integrate big data analytics into their processes were investigated from a rather opportunistic view in mind in a situation where there is no challenges or problems related to big data analytics. Only the good sides of big data analytics usage were handled on the side at the same time as the processes of supplier selection and risk management in the case companies were investigated. However this is not the case when it comes to big data and starting to utilize big data

analytics in companies business processes. There are many restrictions and challenges still related to that. Still interviewees from the case companies see that there is a lot of potential to use big data in business. The additional information and value big data could offer is undeniably an opportunity but first case companies should know how to utilize it and tackle the challenges preventing big data analytics usage. Companies should always bear in mind that big data analytics is an opportunity but also a challenge. These themes will be next discussed more deeply in the following chapters of this empirical part.

7.3.1. Definition of big data in case companies

All the three interviewees from supply chain department in the company Alpha described big data as something that is very abstract as a concept. According to their opinion big data is something that "exist somewhere" in massive quantities and by combining data from different sources you can gain information that is not available from current traditional data sources. From company Beta the director of product sourcing defined big data as all kind of transactions and numbers which you can take out from the system. On the other hand purchasing manager in product and services from USA office in company Beta defined big data as unstructured gigabytes of data that are hard to handle.

From company Beta three interviewees were from digitalization and data management departments but even though they are kind of more experts on the big data concept they still have quite similar definitions for big data as interviewees from supply chain department. Project manager in digital solutions platform and operational excellence from company Beta defines big data as all the data that for example factory has such as process data, measurement data and order data. According to him big data is in a way all data that is related to some question someone is trying to find an answer to. Company Beta's manager in smelting automation digitalization describes big data as data that can be combined from many different sources such as social media. He sees that big data is all kind of data available in the world and there is much of it. Company Beta's MIS specialist on the other hand defines big data as raw data that is gathered for example from manufacturing factories in the form of process data and from all kind of other business systems.

As it was already stated in the theory part of this thesis, there is no one unified or agreed definition of big data as a concept. There are some similarities in the answers

of the interviewees about big data definition but in the most parts they are really varied. Even the employees working more closely with big data related activities did not have that specific definition for big data to give. Defining big data is something that case companies should focus on if they start utilizing big data in their business operations. They need to define big data in a way that is unified and means same thing to everyone within the company. The definition of big data should start from the point of view what it means to the case company itself, so it could be different in case company Alpha than in case company Beta.

7.3.2. Big data related challenges in case companies

As mentioned already in the theory part even defining big data is challenging let alone starting to utilize it in the companies' business operations and processes. Interviewees from company Alpha's supply chain department all agreed on that there are many obstacles and challenges related to big data as a concept. In their opinion big data is still very abstract as a concept and it is not clear at all how to get concrete benefits from using big data in their business operations. The cause and effect relationship between using big data and gaining actual benefits with big data so is still too vague. They state that there has been a lot of data existing in the world even before the term of big data but the main problem that remains in their opinion is how to effectively to exploit the data. Also company Beta's purchasing manager in product and services in USA also states that the problem is that they do not have the full knowledge of big data. They do not know how to use it and what are the benefits if they use it. In turn company Beta's manager in smelting automation and digitalization sees they do not have any big data in company Beta, the data they have is rather small data. He also feels that they do not even know in company Beta how and where to use big data. First it would be first more beneficial to learn how to effectively use small data they have to make better and more durable products and to enhance their on-time deliveries to their customers. Not really knowing the possible benefits of big data usage is a really profound problem which the case companies need to tackle before anything else.

Interviewees from company Alpha also state that there is and will be a lot of challenges in determining how reliable the data is and how the data quality and validity can be ensured and estimated to prevent it from giving wrong signals to the company decision makers. Company Alpha has even had one consult that estimated their data and big data usage possibilities but at that time they felt that big data was not providing any concrete information. They felt that is was impossible to modify big data into valid

information they could utilize. In an optimal situation it should be possible to “squeeze” the good data out of all the big data and summarize it into information that is easy to use as it is. Also according to the director of product sourcing and supply from company Beta a challenge for them is how to get relevant and reliable data for example from their CRM system. Another problem or challenge is also how and where to save all the data and who takes care of the data. Other challenges company Alpha interviewees mention are how to feed the data and the information to the systems and at the same time from many different sources.

Also the interviewees from both case companies stated on many occasions that they do not want to be one of the first movers in big data usage. None of any other companies in the market utilizes big data in their business operations as a tool to help in complicated decision making. There are not that many success stories in big data usage in business operations yet. No company wants to necessarily be the first mover, naturally they would first like to see some real success stories of big data usage in business processes. Also both companies employees mention that the barrier for even to start thinking about big data in their operations is that the benefits are not clear to the customers and to the case companies themselves. They do not even know what big data is to them. Both case companies should first investigate the barriers they have towards the big data and try to fix them the best they can.

7.4. Integrating big data analytics to supplier selection risk review in case companies

Starting big data usage would be somewhat beneficial for both case companies in an optimal situation without any big data related challenges handled in the previous chapter. For instance company Alpha interviewees all mention that they have really close relationships with their suppliers that are most important to them and they are somewhat dependent on. However there is not enough time to have good relationship with all of their suppliers and it would not even be possible or beneficial for the case companies to do so. Still there could be a huge possibility to start utilizing big data analytics in companies’ business operations with their most important and closest suppliers with whom they already have a lot of collaboration activities. At least it would be the easiest way to test and start big data utilization with those suppliers as some of these suppliers are also very dependent on company Alpha so it would be easier to make also them see and comprehend the possible benefits from exploiting big data analytics in their processes.

7.4.1. Case companies' readiness stage for starting big data analytics usage

Company Alpha

According to company Alpha employees the most significant big data channels or sources for them could be everything they can exploit in their business operations. However it is only a statement as at the moment company Alpha does not have any big data related technologies in usage. The interviewees also state that they do not share their information and data with their suppliers as much as they could with their current tools. They do share some drawings with their suppliers with E-parcel tool to which the both parties possess an access to. They are currently trying to develop new tools for information sharing in company Alpha but it is always a slow process and it is not even certain the new developed tools will be used in the end at all. The main issue about information sharing is to get benefits from it and at the same time not to share too much IPR information and other knowledge outside the company borders. Data protection and minimizing information sharing risks are challenging issues for any company but these things needs to be taken into consideration when sharing information with the suppliers.

The general opinion within company Alpha interviewees is that they do not have the needed willingness or enthusiasm to start any big data related business activity but they do have the requirements and skills they would need for integrating big data analytics to their processes. In this sense they are pretty confident and arrogant at company Alpha that it is no big deal to start big data analytics usage. Rather this probably is an indication that big data is not fully and deeply understood in company Alpha. Company Alpha is also at some stage interested in a Finnish company called "Jakamo" which company Beta is currently already utilizing with their suppliers. Company Alpha interviewees think the ideology behind Jakamo is good and that the platform works well. Through Jakamo companies are able to share their documents and manufacturing drawings with their suppliers in real time and at least that is something company Alpha is interested in to start utilizing in their business operations and processes. This could be the very first step for company Alpha when it comes to information and data sharing to be able to start big data analytics usage in their business.

Company Beta

The most important big data sources for company Beta according to the project manager in digital solutions platform and operational excellence would be process data, process history data, all of their databases, different system reports and ERP data. The director of product sourcing and supply mentions as important big data channels and sources for company Beta SAP and Microsoft Dynamics (CRM). He sees that spend reporting his team is using from OBI is one used big data related technology in their department. Company Beta's manager in smelting automation digitalization determines the most important big data sources and channels in the same way as company Alpha employees, it would be everything possible company Beta could benefit and have access on. He mentions at least Matlab as one big data related technology in usage. MIS and MES data management specialist from company Beta stated that the most important big data source or channel for company Beta would be the automation systems in their factories and process data that they get in real time from the different sites they have around the world. According to him they have tested some big data related technologies in their department such as machine learning and proof of concept but they are not really using them in their business operations. On the contrary the purchasing manager in product and services in USA from company Beta feels they do not have any possible big data source or any big data related technologies in use.

According to the project manager in digital solutions platform and operational excellence company Beta does not have an own analyst who could truly analyze big data. He feels that they do not necessarily have the requirements and resources in company Beta that the big data usage requires. He feels that they have the general understanding on big data but what they are missing is the big data analysts and experts. Thus they do not possess the in-depth ability to extract valuable information from big data. They do have few analysts working in the company but that is not enough. Also company Beta's MIS and MES data management specialist feels that company Beta does not have the capabilities internally for big data usage to be able to analyze data for instance through machine learning. They have rather had only external project leaders who have brought the know-how and coding skills in house.

According to manager in smelting automation digitalization in company Beta they are planning to start utilizing a platform where data is shared with suppliers and customers. According to the director of product sourcing and supply from company Beta at the moment they are not sharing that much information with their suppliers. However he

feels that there are definitely possibilities if they would develop this with their key suppliers. It could better their forecasting and they could gain benefits for their capacity planning. Starting to share information more with their suppliers could be a big step forward for company Beta as they are somewhat dependent on their key supplier performance and typically they already have close relationship with them. Thus starting to share information would not necessarily be that difficult. However the director of product sourcing and supply feels that company Beta is also lacking the technologies and skills to start utilizing big data. He feels they only have the open attitude for big data usage. Also according to the company Beta's purchasing manager in product and services in USA states that they do not share information and data within their supply chain. They have good relationship with the suppliers they are dependent on but still they provide limited visibility to them.

From the case company interview answers it can be seen that those interviewees who work more closely with topics and departments related to big data have a lot more realistic view of the case companies' readiness stage for big data analytics usage. Generally it can be seen that the interviewees working in supply departments do not have as wide views possibly because they do not think things like this in their daily work. However most of the interviewees even from the supply departments understand well the possible data sources and channels from which they can gather data and information to exploit it in their business. Still training throughout the case companies would be very crucial if the companies would take big data as part of their operations. With training it would be ensured that everyone in the case companies understand the possible advantages of big data utilization. It is not enough that the employees working with big data understand them as the business sales people need to understand them too so they would carry the understanding towards customers too. The first step for both case companies could be opening the data and information sharing with their suppliers and other partners to get closer to starting to use big data analytics.

7.4.2. Challenges of integrating big data analytics to supplier selection risk review in case companies

Previously in this empirical part it was already talked about the challenges of big data as a concept. This chapter discusses big data challenges more from the supply side in a situation when and if companies start to use it in their processes. The challenges handled here are more from the point of view what kind of challenges utilization of big data causes. As it was already commented earlier in this empirical part about big data,

there exists a great myriad of challenges with the concept of big data alone. When companies proceed further to integrate big data to their supplier selection, the challenges and problems naturally multiply. Most of the challenges related to integrating big data to supplier selection is from case company Beta's point of view as they had a bit more to offer on the topic mainly because some of the employees were from the departments working with analytics side and had more working experience on the topic.

"Big data to us is still like guns to a person in Egyptian time. Asking why we do not utilize big data in our business operations is like asking a person in Egyptian time why they do not use guns. The answer is simply because the bow and arrow are the only things they know how to use." (Purchasing Manager in Project & Services (in USA market), company Beta, 2018)

Project manager in digital solutions platform and operational excellence from company Beta thinks that starting to share data and information within supply chain is a great challenge as no one shares their data and information willingly. Further it is hard to get any information especially from small suppliers who have little or no information public or otherwise available. Also if the project is small it might not be profitable to use big data in that process at all. In addition he thinks that usually supplier selection is constituted of many small things in most of the companies so he feels it might be difficult to unite supplier selection and big data. MIS and MES data management specialist in the company Beta shares the same views. The situation needs to be win-win situation for both the company Beta and the customer but also for the suppliers for all parties to agree on sharing data and exploiting the shared data in their business operations. They also feel in company Alpha that it might be hard to get information on suppliers as is usually costs and the systems and tools for it also costs and nobody wants to pay if the benefits of sharing data are not really concrete and visible.

Data ownership is a big challenge as no one really wants to share their data. The project manager in digital solutions platform and operational excellence noted that the main problem and obstacle for company Beta is that they do not own the data they should be exploiting to really benefit from big data usage. The most beneficial data is mostly customers' data and it is very hard to get their hands into that data. There is a lot of bureaucracy and expensive contracts preventing company Beta from assessing their customers' data even in small amounts. Also in order for customers to even agree to give out their data to company Beta's usage they would first need to see and understand the concrete benefits they could get from data sharing. Still also he sees

big data usage as an opportunity to make their processes better and more efficient. He feels that his business department could offer a lot of big data related services and products that could contain useful data for business purposes of company Beta.

Manager in smelting automation digitalization in company Beta notes that it is also hard to get someone on board who has done their job in the same way for decades. He feels company Beta in a sense has the knowledge of big data but they do not have the willingness to start utilizing it as the company culture is rather old-fashioned. Many employees in company Beta are used to rely on old history knowledge and know-how. They have the mentality of “this has been always done like this so why change it”. There would most probably be some resistance within the employees so it would be very useful to make everyone in company Beta to understand the concrete benefits big data usage can offer. Also the industry in which company Beta operates on is very conservative. All the parties involved have the fear of what data they can share outside the company borders. Also he notes that employees in company Beta are too used to using Excel to handle data but big data does not work in Excel as it is too massive. What is more according to company Beta’s MIS and MES data management specialist not only the company Beta itself but also their customers are very conservative. Either is not that interested on new trends such as big data. The customers are not keen on sharing and exploiting their manufacturing and process data. Overall the whole industry company Beta but also company Alpha are operating is a very conservative one.

From the company Alpha’s point of view big data usage is also challenging as internally their information is in 20 different hubs across the world. Gathering the internal company data together is a massive challenge to them let alone gathering the external data together. They hope in the future there could be one program that could communicate well with different systems to find and integrate information from many separate systems to the usage of company but it is very difficult to make this kind of program. Also company Alpha interviewees note that one challenge with big data usage in business operations could be that there are different ways of doing things inside the company and its subsidiaries. Companies should first have a united way of doing things with big data and information and data sharing to make it more easier to start utilizing big data in their business operations.

What can be noted from the interview answers about the challenges of possible big data utilization in the case companies is that they are almost as varied as the answers about the definition of big data. Even employees working in the same case company have different opinions of the big data usage challenges related to their company. This

proofs that there has not been any or very little discussion about starting to exploit big data in their business processes. For example in company Beta most interviewees had very strong opinions that company Beta does not have what it takes to start big data related activities but the director of product sourcing and supply felt there is really nothing preventing company Beta from it resource wise. This indicates that the interviewees not working within supply department had a bit more realistic and deeper opinion about the challenges than the interviewees working in the supply side of the case companies. Case companies also should select judiciously the situations in which they start to exploit big data and not to invest and integrate big data into everything they do but rather test it first with some operations.

7.4.3. Changes needed in case companies for big data analytics usage

Project manager in digital solutions platform and operational excellence from company Beta states that first they would have to figure out a way to get customers to give out their data for company Beta. That is why it would be important that the sales people in company Beta would understand the big data usage and information sharing benefits and sell that understanding to the customers. However before this mode of operation the vision for this would have to come from the company Beta's senior management and then get the big data analysts on board. MIS and MES data management specialist in the company Beta points out that starting to utilize big data in their business operations is a big investment moneywise. He thinks company Beta should first have one big data department who has the know-how of big data and they would participate as experts to different projects. He feels that company Beta should not establish any major big data department and not use too much of their resources to it. They should first have internal skills related to big data in company Beta. There is also a need to train different product lines within company Beta as both product lines and sales people need to see the big data benefits as their support is also needed as without it the successful usage of big data would be hard.

Director of product sourcing and supply states that company Beta needs to start from the very beginning with big data utilization. He thinks big data is not mentioned anywhere so company Beta should first define as a company what they want. However he also sees that there are not immediate possibilities for selecting new suppliers and rather feels there are more opportunities when working with existing key suppliers. Manager in smelting automation digitalization feels that company Beta should strive to have same approach and attitude towards big data throughout the company. They

should first have some concentrated smaller group who starts deploying big data in company Beta's business operations little by little. Then they could have one person in every business line that could be the so called big data responsible and slowly integrate big data to the whole organization. However for this company Beta would need to have the right tools, software and the right personnel who can handle big data and select the right service providers. He feels that company Beta should first start with small data.

Interviewees from company Alpha felt that possible big data usage would not affect their processes or organization that much. They feel that big data would be kind of auxiliary activity next to old processes. What they feel big data would change is that there would be more checking points in their processes that would prevent the decision makers from making wrong decisions. However in their opinion company Alpha would have to harmonize their nomenclature they use in all of their business operations and in some areas they might need to harmonize the whole process to be able to gather information effectively and to be able to utilize big data. Thus harmonizing the whole process is a huge task to complete. Company Alpha interviewees feel that they would need to be able to integrate big data from different sources in real-time and bring certain parameters and attributes to the process that are related to the supply chain to get up to date information for decision making. They mention they could for example monitor and collect information whether the suppliers have delivered on time, what information there are in general available about the suppliers and claims made of the supplier. All in all company Alpha interviewees' opinion is that they have the knowledge, the skills and the resources big data utilization requires. They state that only their process and way of doing things would change a bit as they would need to add some big data checking points to their business processes. Rather what company Alpha would need to modify and change is their mindset to make big data usage work. However it is a big challenge to make everyone in the company to understand the potential of big data and how big data could truly enhance the business processes.

8. DISCUSSION AND CONCLUSIONS

The main research problem of this thesis was to study how big data can be utilized through risk management process in supplier selection to make the both processes more accurate to select non-risky suppliers to companies' supplier base. The intention was to gain a more comprehensive understanding and picture about the topic as there is not much research related to the topic. The theoretical background handled earlier in this thesis gives a good context for a new empirical case analysis research with two case companies operating in the same industry. Now in this part of the thesis, the existing theory and empirical results are viewed and mirrored against each other to extend the insights of the thesis topic and to discover similarities and differences to build linkages between the theoretical and empirical part. First, the research questions that are based on the main research problem are investigated individually to give answers to them. The research questions are examined according to the theory and empirical findings of this thesis to find similarities and other observations. Then the benefits companies can gain through utilization of big data analytics in the risk review related to supplier selection are presented. After that, the resolution to the main research question and a suggested framework about how companies could start to use big data through supplier risk management process in their supplier selection is given.

This study finds that big data usage for companies is a huge opportunity that brings a lot of challenges that need to be tackled before starting any big data initiative. The main challenge is to change big data into usable information which can be used to audit and select suppliers. There is a big difference between big data and valuable information. It is hard to process big data to get the useful information from big data. Still big data analytics is an opportunity to enhance companies' old-fashioned ways of doing things, to save time and costs, and overall be more efficient in everything companies do. The benefits big data can give to companies' supply management departments and supplier risk reviews are noticed but the full extent of how powerful big data can be in the risk management context is not yet fully realized. However before anything else companies would need to make themselves, and their stakeholders such as customers and suppliers, see the benefits big data usage could bring. This is the biggest and most profound barrier for big data utilization as if no one in the supply chain sees big data analytics as a positive thing and a huge opportunity, nothing will change and big data analytics' positive potential will not be exploited.

8.1. Summary of the research findings

The results of this research are presented here by answering the research sub-questions that help to answer the main research problem of this study. The sub-questions of this study support the answer of the main research question of how big data can be exploited in a risk review related to supplier selection to make the process more effective and supplier selection more accurate to prevent risky suppliers from entering companies' supplier base, so they will be answered first individually before presenting an answer to the main research question through a suggested framework.

What is the supplier selection process (and the criteria behind the selection)?

Supplier selection is very important and crucial function for both case companies and that is good as Rao et al. (2017) state that efficient supply chain management requires central role for the selection process of suppliers. Nowadays the successful supply chain management requires strategic relationship with company's suppliers (Li, Yamagucki and Nagai, 2007). The supplier relationships are rather cooperative than adversarial. That is why having and selecting a reliable suppliers is very important since this enables companies to have their full attention and focus on their main goals. (Mohtasham et al., 2016) Also Mavi et al. (2016) emphasize the fact that nowadays markets are highly competitive, and companies become more and more dependent on their suppliers. Therefore companies have to understand the importance of supplier selection. Supplier selection is a key element in procurement, but it is not an easy task. (Mavi et al., 2016)

Usually researches offer a step by step model for supplier selection and one of those is introduced in the theoretical part of this thesis, but in the practical real life this is not always the case that the selection process has exact steps and model companies follow. According to different situation and needs, the supplier selection process can vary a lot. Some of the interviewees had same views of the supplier selection process steps as Sonmez (2006) whose model of supplier selection was handled in the theory part. According to Sonmez (2006) the supplier selection steps are the following: identification of a need for supplier, determination of decision criteria, pre-qualification (selecting few suppliers from a large list), final supplier selection and lastly continuous monitoring and assessment of supplier. Both case companies' interviewees state that they do not utilize any specific named theoretical model for selecting suppliers, but still

they have rather strict and clear process steps they follow for supplier selection at least when they select suppliers to their supplier base. Even though many of the interviewees from both case companies state that they do not have strict model for supplier selection, they still follow approximately and roughly the same supplier selection steps as Sonmez (2006) has stated in his research when they select suppliers to their supplier base. However when case companies select suppliers for their specific projects from their supplier base neither case company interviewees' named any specific process for project-specific supplier selection as the process might vary a lot depending on the situation and project. So supplier selection process steps vary depending on the situation.

There is also a very vast amount of different supplier performance metrics and supplier selection criteria that are introduced in the theoretical part. Companies need to identify and select those supplier selection criteria that are most important for them. Companies should investigate both commercial criteria and different internal and external supplier risks when selecting suppliers to their supplier base. The most important supplier selection criteria company Alpha has is the supplier's technical competence. Suppliers must be able to manufacture the company's products at the right quality level. All in all the suppliers must have good quality products and be suitable in size and price. Company Beta has some same supplier selection criteria as company Alpha. The most important supplier selection criteria for company Beta is supplier's technical suitability including quality, competitiveness (price) and supplier's interest to work closely with company Beta.

Both case companies also use many sources to gather information on the suppliers when selecting them and with big data analytics they could combine these multiple sources into one to get the all information available about the suppliers. Further big data analytics could help case companies with the challenges they encounter with the selection process such as the problem of not getting relevant or accurate information as the supplier selection and especially global supplier selection is a choice with high degree of uncertainties, risks and fuzziness. Further supplier selection is a very complicated multi-criteria and multi-objective decision. With BDA it could be easier to make the decision that needs many aspects to be considered when making the decision. However there are some differing answers regarding the supplier selection process and information gathering practices among the interviewees even working in the same company. This could be due to the fact that some of the interviewees were working in different subsidiaries of the case company so they might have adapted a different ways for supplier selection in the subsidiaries of the case companies.

What is the risk review or management of suppliers in the supplier selection process (and the risk criteria that must be assessed)?

There are a lot of risks existing related to suppliers themselves but also to the suppliers' locations and markets they are operating in. These are for example supply risks, operational risks, demand risks and security risks (Manuj and Mentzer, 2008). What needs to be noted and accepted is that uncertainty and risks are very present in the supplier selection process (Patra and Mondal, 2015). Trkman and McCormack (2009) emphasize that is especially important to recognize the suppliers that have an increased potential for a disruption to prevent them from even entering the supplier base in the first place. Companies must constantly measure their supplier performance to be able to predict and manage disruptions and risks (Trkman and McCormack, 2009). All the possible risks related to one specific supplier must be investigated and estimated to be able to be as prepared as possible if the risks are realized but it is not a simple task for companies.

According to Mensah et al. (2017) modern supply chains are vulnerable to much bigger risks than managers of supply chains are even able to recognize anymore. The amount and types of risks that supply chains encounter are now greater than before and that is why the risk management process has become more complex and at the same time more important than before. (Mensah et al., 2017) With the help of big data analytics the whole risk management process could be made a lot easier and accurate as the risks are always hard to forecast and are often cyclical. However everything cannot be controlled such as the climate and weather but with big data analytics companies could get early warning signs of possible problems before they are realized. With big data analytics it is also possible to achieve the situation where the data utilized is actually up-to-date which itself would make the risk management process all in all more accurate.

The risk management process steps recognized in the theory part of this thesis are the following: risk identification, risks assessment and measurement, risk management decisions and finally risk monitoring and control. Same as with the supplier selection process steps, in companies there are not necessarily one specific theoretical method for risk management in supplier selection. All the company Alpha interviewees state that they do have some risk management process steps in place but they are not that specific. However they do use many different tools and systems and even utilize the services of an external risk method company for risk management but their risk

management process is still not that systematic. Company Beta's interviewees state that they manage and assess the supplier risks already when they approve the suppliers to their supplier base. Suppliers are screened for technical, financial and EHS capabilities before approving them to the supplier base. Company Beta also has a separate project specific risk management process in place which covers the supply chain and suppliers. This means risk management is not done on a company level even though they have many tools for risk management in usage.

Company Alpha interviewees feel that there is no specific supplier risk that is greater than any other risk as anything can go wrong. The most common risks they have encountered are supplier quality risk, delivery and logistics risks, and "unexpected" capacity risks of suppliers such as bankruptcy and other finance risks but there are also country risks such as political risks in the supplier's markets. For company Beta quite typical risks are related to supplier capacity constraints and the most severe risks are quality risks but also poor delivery. They also mention as company Alpha interviewees that their and their suppliers' markets cause risks for them. For both case companies the person in charge of risk management is almost always project specific. In every project company Alpha has they assign a certain person to look after risk management so in neither company the risk management is not done on the whole company level. In company Beta all of the project members are responsible for the risk management, but also still the project manager has the ultimate responsibility of the risk management process. As the case companies do not have any specific systematic and comprehensive way for doing risk management, it could be easy to adapt the process to be more suitable for big data analytics and then change the process be more systematic and precise.

What is big data?

The results from the empirical part largely supported the findings of theoretical part that big data as a concept does not have a one unified definition and that it can mean many different things to different people. Different authors have varying definitions for big data that were introduced in the theoretical part. It also means different things to different people working in different contexts and departments and this can be seen from the answers interviewees gave when they were asked the question what is big data and how they would determine the concept. The interviewees defined the term on a more general level rather than being able or wanting to explain it in more depths from different ankles. None of them for instance knew or at least did not mention the basic characteristics of big data described in the theory part (volume, velocity, variety,

veracity, value, visualization and variability). After all the characteristics are the features of big data that distinguish it from other data and what are the basic source to the big data related challenges. Enterprises need to jointly and comprehensively strive to handle big data while taking all of its characteristics into account (Emani et al., 2015).

Some of the interviewees were able to qualify some possible big data sources and big data related technologies but not in great extend as neither case company is not yet that encaged to starting to utilize big data. Many of the interviewees did not quite comprehend all the data source possibilities they have and are exposed to as especially the global supply chain management industry is having a huge and increasing big data information amount that is flooding constantly from different sources in real time such as sensor networks, digital machines, and mobile equipment. However of course it needs to be noted that all this data is useless if the companies do not know how to use it. In addition none of the interviewees mentioned for example the concepts of parallel computing or cloud computing that were discussed in the theoretical part of this thesis. However machine learning was mentioned. This is probably due to the fact that the case companies have not looked into the topic as they are not yet ready to start integrating big data analytics to their business processes, so they are not familiar with all the concepts.

One interviewee from company Beta described the very essence of big data opportunity: big data is in a way all data that is related to some question someone is trying to find an answer to. This is very important remark as it defines the basic benefit of big data that companies can obtain by using it: with big data it is possible to have access to all important data that is related to some problem or question they have. Big data truly gives supply chains decision-making a greater data clarity, accuracy and new insights that can lead to better intelligence across the supply chain - big data can be the intense force for driving supply chains ahead (Tiwari et al., 2018). However companies should first strive to determine big data from the view point what it means to them and what they want to achieve with it. After that the case companies should share this with everyone within the company and ensure the definition means same thing for everyone. The definition of big data in case companies should start from the point of view what it means to the case company itself as it might mean a very different thing for company Beta than for company Alpha depending on what the company wants to achieve with big data and where they want to use it in their processes and what are their objectives.

What companies need to have to be able to utilize big data and what are the challenges of using big data?

There is a great myriad of challenges related to big data and especially in the utilization of it in the business processes. All of these challenges should be identified and analyzed before starting any big data activity in the case companies as they could generate resistance towards big data within the employees if they are not first handled. In the theory part it was found that there are many opinions between different authors about which are the most significant big data related challenges and this seems to be the case also among the interviewees from the case companies. In the theoretical part the challenges were categorized under wider categories according to Sivarajah et al. (2017): data characteristics challenges, process challenges and managerial challenges. The main challenge is to analyze big data in a way that brings value. (Sivarajah et al., 2017) Other main challenges mentioned in the theory part are for example the following ones: data and intellectual property protection and security, lack of knowledge and skills, lack of access to expertise, budget restrictions, technical challenges and data ownership.

The challenges mentioned by the interviewees are for instance the following ones. Big data is still too abstract as a concept and it is expensive to start big data analytics operations. The cause and effect relationship between using big data and gaining actual benefits with big data so is still too vague. Many of the interviewees also mention that they do not know how to change big data into useful information and if they can trust the data is reliable and valid. Also data gathering, storage and management problems of big data were mentioned in the interviewee answers and the fact that no one wants to share their data and information willingly. Further case companies' subsidiaries might have very different ways of doing things which can be a bit problematic as companies should first have a united way of doing things with big data and information and data sharing to make it easier to start utilizing big data in their business operations. Data ownership is also a massive obstacle according to the interviewees. What can be noted from the empirical findings is that interviewees working in departments of digitalization and such departments had wider views on this topic as their work probably is already somewhat related to big data, at least more than those interviewees' work that were from the case companies' supply department.

General opinion in case companies Alpha and Beta seems to be that they feel they have the requirements on big data usage but they do not have the willingness to start utilizing big data in their business operations. Few of the interviewees in company Beta

from digitalization department stated they do not necessarily yet have what it takes to use big data. Neither of the case companies share their information with their suppliers at least not in a great extent. They are rather only curious about the possibilities to share information with different tools but only if it is safe. They do not want to be the first movers as in their opinion there is not that many success stories related to big data usage and it is not that common at all to utilize excessively big data in business context. They feel they would be more interested if there would already be more success stories related to big data usage. Also both of the case companies and their industries are rather old-fashioned and conservative – they are very keen on doing things like they have always been done them.

However to be able to tackle the above handled challenges related to big data companies need to have certain skills, knowledge and technologies in place. According to Militaru et al. (2015) supply chain partners seeking to utilize big data analytics need more than just novel skills, technologies and tools. In fact they have to think through the way they function and alter their business processes and mindset. The enterprises can benefit from implementation of big data only by innovating in processes and operations. To be able to use big data in the companies' business processes, the organizational and technological changes have to be implemented across the whole supply chain to ensure information sharing is possible between the supply chain partners. (Militaru et al., 2015) At the same time starting to use big data is a big investment to get the all needed resources to make it work. Big data usage is a possibility, but the challenge remains how to get everything they need to start using it. They would also have to change their company culture and basic perceptions a lot. Case companies need a data-oriented attitude because now the case companies rather rely on old history knowledge and expertise. They are very keen on doing things the way they have always been done and this need to be changed. However first and foremost they would need to make everyone see the benefits of big data because without that none of the stakeholders in the companies' supply chains will agree to share information.

What big data related methods and techniques are there that can be utilized for supplier selection risk review?

Big data has already enhanced a lot the development of risk management solutions. Big data can enhance the risk management models' quality because there is an increasing diversity and availability of statistics when utilizing big data. Big data can be used to simulate a range of scenarios to be able to recognize all of the potential risks.

This leads to faster reactions to changes in the market. Big data helps to identify hazards faster by comparing external and internal data companies have. For instance social media analysis can be used for risk management. (Meraglim, 2017) Companies are always in the end responsible for what happens in their supply chain that is why they have to select the suppliers carefully and understand the risks behind every supplier. To do this, companies need to integrate their internal supplier data with the huge amounts of third party data or unstructured data available to develop deeper understanding of their supply chain as a whole. Companies need to profile their supplier base against third party data to avoid supplier risks. For instance political and geographical mapping are important methods to do this. Especially global supply chains have to understand other nations from political, financial and geographical point of view. Big data is the answer to understanding these risks. By complementing internal data sources with external data companies can start to understand what is happening outside their company boundaries. Incomplete supplier information is a huge risk for companies. It is impossible to totally delete supply chain risks but the visibility given by big data makes it possible to better handle the consequences of the risks.

None of the interviewees were able to name that many actual big data analytics related methods, technologies or techniques that can be used in business processes and this is simply because the case companies have none in usage in their business operations. For instance Matlab and machine learning were mentioned only by one of the interviewee working within digitalization. However there are some methods or techniques that can be used to integrate big data analytics into supply risk management and supplier selection process and those are handled more extensively in the theoretical part of this thesis. The ones mentioned in the theoretical part are for instance predictive, descriptive and prescriptive analytics, statistical analysis and optimization, modeling and simulation and visualization. Also social media and public news data (social media analytics) can be utilized as one big data source and analysis method. By altering the different methods to the companies needs these different methods could be a part of the companies supplier selection and risk management processes.

With big data it could be easier to connect different supplier selection criteria together to make the decision-making information available more comprehensive as the supplier selection decision is never based on a single evaluation criterion. It is a multi-objective and multi-criteria decision. It is a combination of many things and with big data analytics it could be possible and a lot easier to connect all this information from different sources to make the supplier selection more accurate and at the same time to

lessen the risks from the supply chain as a whole. Big data analytics could be used to combine company's selected supplier criteria, and internal and external supplier risks and other data available that affect the supplier selection decision to gather all the information needed for the selection process to make it more accurate for companies.

8.2. Benefits of big data analytics usage in supplier risk review related to supplier selection

Carefully done supplier selection is one of the most significant decisions to be made in supply chain management as it has a massive effect on companies' cost reduction and competitive advantage. (Lamba and Singh, 2018). However it needs to be noted that balancing the costs of big data analytics usage and its potential benefits are challenging for any enterprise. Wang and Alexander (2015) made in their research a list of benefits companies gain when using big data analytics in supply chain management. Those benefits are categorized in four categories: accuracy of information and increased visibility, higher service quality, operational efficiencies, and better prediction and forecasting. Visibility and information accuracy makes it possible to notice risky suppliers and possible risks suppliers might encounter in the future by giving early warning signs of potential supply chain risks with the help of big data analytics. Gaining operational efficiencies means faster and more accurate supplier selection and supplier management in real time. At the same time the usage of big data analytics improves the companies' overall service quality through close collaboration with the supply chain partners. (Wang and Alexander, 2015) The forecasting and decision making becomes more precise with the help of big data analytics because with big data analytics the decision and supplier selection is based on evidence and a massive amount of information provided by big data analytics.

All in all big data analytics is a potential solution for improving supplier risk management and selection as risk management is usually very difficult to manage thoroughly. Data sharing and big data analytics usage lowers the decision-making risks and thus decreases the possibility of selecting a risky supplier to company's supplier base. There lies a lot of risks in companies' suppliers and big data analytics can aid in managing them more effectively and comprehensively. What is more is that supplier selection is multi-objective and multi-criteria decision and both case companies need to use many different sources to gather all the possible information and data on the suppliers when selecting them. Big data analytics could combine these multiple sources and criteria into one to get the all information available to make the correct

supplier selection. Further big data analytics utilization could fix the problem companies usually encounter with information retrieval; the information from different sources is rarely up-to-date and many random things affect the decision making such as weather. It is hard to combine the selection criteria and all the information from all the possible data sources to get the all information available for supplier selection. Big data and BDA could help with these issues and make the decision making and forecasting more accurate to ensure the correct supplier selection to companies' supplier base.

8.3. Framework for starting to use big data analytics in companies' supplier selection risk review

Either of the case companies is not yet utilizing big data analytics in any way in their business processes. They have rather only been curious about the topic and investigated the big data possibilities a bit, but nothing really concrete has happened in the case companies towards starting to actually utilize big data analytics in their supplier risk review and supplier selection processes even though to some extend the both case companies understand the possibilities and benefits of big data anaytics. In this section the thesis suggests a framework or guidelines how companies could possibly prepare themselves to be readier for big data utilization or how they could start to test big data analytics usage in their process of risk review related to supplier selection (Figure 11). The basic core behind the framework is the basic risk management process, where risks are first identified and analysed, then evaluated and treated accordingly (Tummala and Schoenherr, 2011). The correct supplier selection to succeed also demands constant communication, consultation, supplier information gathering, monitoring and reviewing of suppliers as constant side processes.

To use big data analytics with success, it is important to have a structured evolutionary approach to accommodate the big data's broad scope. Firstly, companies should start to use their internal data to have a clearer picture of data sources that would benefit the company so the company can go and expand to external sources. What is more important than the data amount is to have an integrated data analysis process. Taking small steps in implementing big data systems allow companies to identify areas of weaknesses and risks. (Meraglim, 2017) Companies should also select judiciously the situations in which they start to exploit big data and not to invest and integrate big data into everything they do but rather test it first with some operations that are most suitable for the big data analytics usage and would benefit most of it. Further there is a great myriad of challenges related to big data and especially in the utilization of it in the

business processes. All of these challenges should be identified and analyzed before starting any big data activity in the case companies as they could generate resistance towards big data analytics within the employees if they are not first handled.

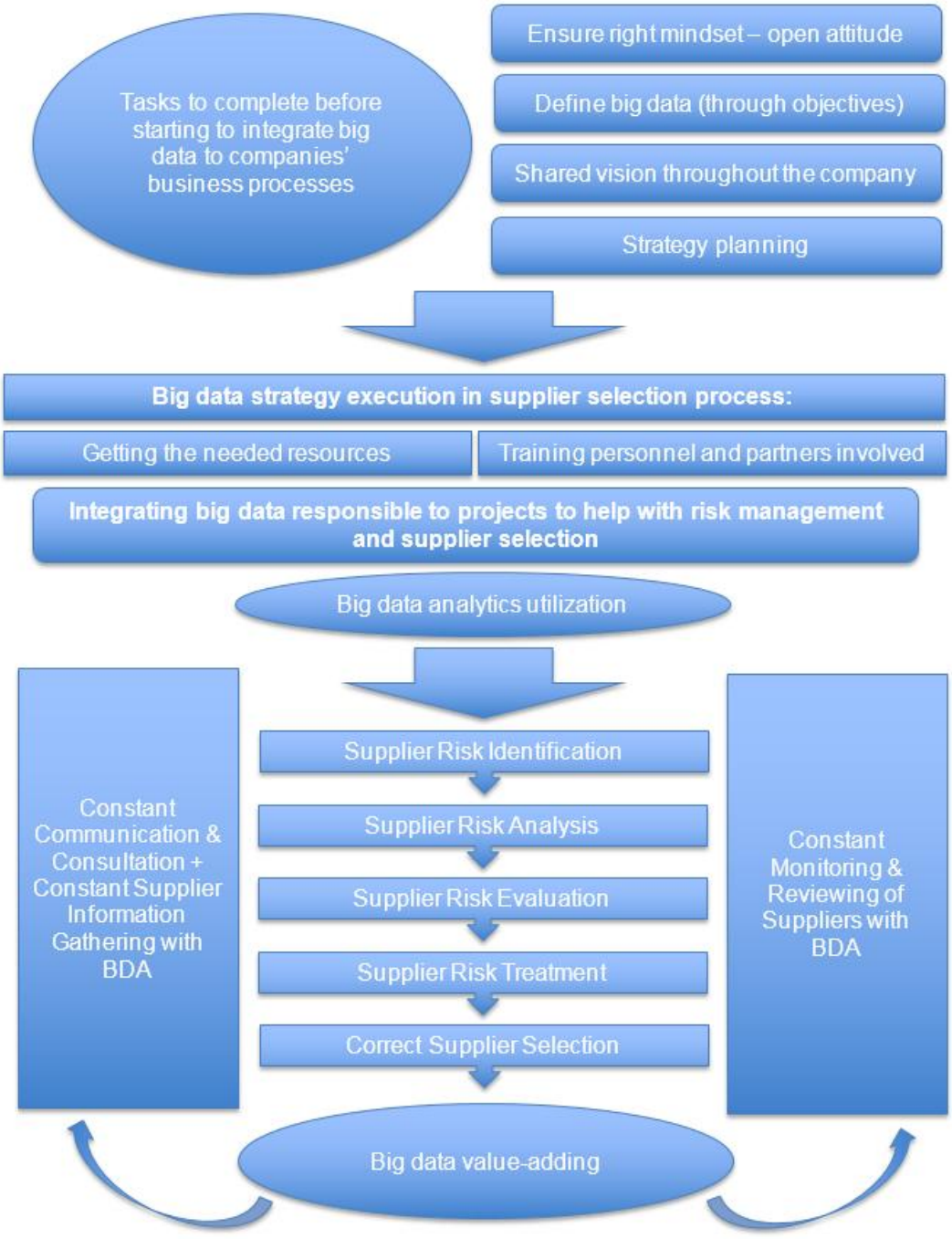

Figure 11. Framework for starting to use big data in companies' supplier selection and risk management processes

First and foremost, it is important to have a vision in the companies for example that big data will be made into a competitive advantage and make concrete objectives for big data usage. The companies need to change their mindset they have about big data

and strive to be more data-oriented and have an open attitude towards big data utilization. The vision should be concrete, limited in time, understandable, realistic and inspirational for it to work. Companies should have a desire to use all the accessible data and to see this as a possibility and not as a threat. What is also important is to determine big data from the view point of what it means to the company, what the company wants to achieve with it and what are the companies' objectives of big data usage. The benefits of big data usage need to be made visible and concrete to everyone in the company's supply chain – the company itself, its customers and suppliers.

The next step after achieving the shared vision and attitude towards big data throughout the company would be strategy planning and execution based on the objectives set. However strategy is something that many enterprises lack when it comes to big data utilization. Still planning the strategy how to start big data related operations in their organizations' processes is a very important step for companies planning big data analytics utilization. The whole company should take a part in big data strategy. Also for the implementation to work the process should start from the top to the bottom – the top management needs to be committed to the change. To help in this companies could have consults or other big data experts on board who would explain the benefits, possibilities and provide training. Companies also need to make room for big data analytics processes to ensure the open information sharing that is needed for big data analytics usage to succeed. What is also needed is all the needed resources for big data analytics usage and training for employees.

One example to start big data operations is to first have some concentrated smaller group who starts investigating and deploying big data analytics usage in companies' business operations little by little. This group would have the responsibility to gather all the needed resources companies need for big data analytics usage and provide training to employees. Later these people would also be the ultimate big data responsible people in the company's processes utilizing big data. Companies could have one person in every project that could be the so called big data responsible who would help in getting needed information for risk management and supplier selection process. After that companies could slowly integrate big data analytics usage to the whole organization if needed and seen even necessary. What needs to be noted is that companies might have to harmonize their nomenclature they use in all of their business operations and in some areas they might need to harmonize the whole process to be able to gather information effectively and to be able to utilize big data in supplier selection. Thus harmonizing the whole process is a huge task to complete. To

conclude big data analytics utilization brings value to the whole process of supplier risk review related supplier selection and ensures the correct supplier selection.

8.4. Theoretical and managerial contribution

This research contributes to the theoretical discussion of big data utilization in business context as the aim of this study was to examine how big data can be used in risk management process and supplier selection process to improve them and make the supplier selection more accurate. There is also not that much researches related to big data utilization in supply management context at all. Further there is a huge research gap in the existing literature to studying big data, supplier risk management and supplier selection together. This thesis strives to combine the concepts of big data, supplier risk management process and supplier selection together to find out how big data analytics could make supplier selection process more accurate to prevent the selection of high-risk suppliers. There is a lot of opportunities for companies if they start to use big data analytics to develop the processes of supplier selection and risk management further.

Altogether, the empirical results in this study were compatible with the existing theory about the supplier selection process, risk management process and big data. However in some extend the empirical part results differed from the findings in the theoretical part. This likely resulted from the interviewees' different subjective views on the topic and simply from the fact that the interview answers are more realistic and practical from the company point of view than those found when conducting the research for the theoretical part from the existing studies. Further, this research contributes to the discussion of starting to utilize big data in companies processes by providing a framework or suggestion how companies should start to integrate big data into their business operations of supplier selection and risk management. The framework is based on the theoretical part and empirical part findings and offers concrete steps what companies should consider executing if they are interested in starting to use big data analytics in their operations.

8.5. Limitations and directions for further research

In total eight people from case companies were interviewed for this research. Even though data gathered from these interviews is quite comprehensive, researches in the

future could grow the number of interviewees who have more insights, knowledge and experience on big data. Now only three interviews had more deep thoughts on big data as they were working in departments more related to big data than rest of the interviewees working just in supply departments. In that way the research could gain more insights into the big data side. Also what needs to be noted that the interviews conducted via e-mail are not equal in quality to the interviews conducted face to face. Face-to-face interviews produced more in-depth data than the email answers. During the face-to-face interviews it is easier to explain the interview questions and ask extra questions from the interviewees during the conversation to avoid misinterpretations.

Even though qualitative case study is suitable for the objective of this thesis, there stems a limitation from the selected research method in the empirical part. The results from a case study cannot be globally generalized. This study was conducted as a case study that focused on two case companies. The case companies in the empirical part are both from the same business industry so there is already one limitation regarding the empirical part of this thesis. Some other not so conservative industry would be readier for big data usage and the results could be a lot different. Further the point of view is rather from the case companies' perspective. That is why it would be interesting to get the suppliers' or customers' point of view to the study by interviewing also them. This could offer a much deeper understanding to the topic. Also it would be interesting to investigate other business processes in companies than just risk management and supplier selection and how big data could be used in other processes companies have.

The main aim of the research is to examine the companies risk management and supplier selection practices and how big data could be used in these processes to make the supplier selection more efficient. Multiple case study having case companies from different stages of big data usage could also bring more and more comprehensive insights for companies not yet utilizing big data in their business processes. So further qualitative researches could be done from many different viewpoints to find other perspectives and views on the topic investigated. Future research could have more case companies involved and different viewpoints from different industries. That way the research could gain more interesting new vision to the research and perhaps also lead to getting even deeper insights. Further as big data is such a new research area in the context of supply chain management, one future research recommendation is longitudinal study. That would bring valuable information about how the usage of big data in supplier selection develops over time.

REFERENCES

Articles, journals and books

Addo-Tenkorang, R. and Helo, P.T. (2016) Big Data Applications in Operations/Supply-Chain Management: A Literature Review, Computers & Industrial Engineering.

Akbay, S. (2015) How big data applications are revolutionizing decision-making. Business Intelligence Journal, 20, 1, 25-29.

Akerkar, R. (2013) Big data computing. CRC Press.

Amid, A. Ghodsypour, S.H. and O'Brien, C. (2009) A weighted additive fuzzy multiobjective model for the supplier selection problem under price breaks in a supply chain. International Journal of Production Economics, 121, 323-332.

Balaraj, S. (2013) Optimization model for improving supply chain visibility. Infosys Labs Briefings, 11, 1, 9-19.

Beroggi, G. E., and Waliace, W. (2000). Multi-expert operational risk management. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 30, 32-44.

Bikakis, N. (2018) Big data visualization tools. Encyclopedia of big data technologies, ATHENA Research Center.

Blackhurst, J.V., Scheibe, K.P. and Johnson, D.J. (2008) Supplier risk assessment and monitoring for the automotive industry. International Journal of Physical Distribution & Logistics Management, 38, 2, 143-165.

Blau, J. and Gobble, M.M. (2015) Big demand for Big Data scientists in Europe. Research Technology Management, 58, 3, 3.

Bok, B., Caratelli, D., Giannonen, D., Sbordone, D. and Tambalotti, A. (2017) Macroeconomic nowcasting and forecasting with big data. Federal Reserve Bank of New York Staff Reports, no. 830.

Borghesi, A. and Gaudenzi, B. (2013) Risk management: Operational Risk and Supply Chain Risk Management. Verona, Springer-Verlag Italia.

Chai, J., Liu, J.N.K. and Ngai, E.W.T. (2013) Application of decision-making techniques in supplier selection: A systematic review of literature. Expert Systems with Applications, 40, 3872-3885.

Chan, F.T.S. and Kumar, N. (2007) Global supplier development considering risk factors using fuzzy extended AHP-based approach. The International Journal of Management Science, Omega, 35, 417-431.

Chen, C-T., Lin, C-T. and Huang, S-F. (2006) A fuzzy approach for supplier evaluation and selection in supply chain management. International Journal of Production Economics, 102, 289-301.

Chen, H., Chiang, R.H.L. and Storey, V.C. (2012) Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly, 36, 4, 1165-1188.

Chen, J., Tao, Y., Wang, H. and Chen, T. (2015) Big data-based fraud risk management at Alibaba. The Journal of Finance and Data Science, 1, 1, 1-10.

Chen, M., Mao, S., & Liu, Y. (2014) Big data: A survey. Mobile Networks and Applications, 19, 2, 171–209.

Chen, W. and Zou, Y. (2017) An integrated method for supplier selection from the perspective of risk aversion. Applied Soft Computing, 54, 449-455.

Choi, T-M., Chiu, C-H. and Chan, H-K. (2016) Risk management of logistics systems. Transportation Research Part E: Logistics and Transportation Review, 90, 1-6.

Dean, J., Ghemawat, S. (2008) MapReduce: simplified data processing on large clusters. Communications of the ACM, 51, 1, 107–113.

Dekker, R., Pinçe, Ç., Zuidwijk, R., and Jalil, M. N. (2013). On the use of installed base information for spare parts logistics: a review of ideas and industry practice. International Journal of Production Economics, 143(2), 536–545.

Demirkan, H. and Delen, D. (2013) Leveraging the capabilities of service-oriented decision support systems: putting analytics and big data in cloud. Decision Support Systems, 55, 1, 412–421.

Eichengreen, B., and Gupta, P. (2013). The two waves of service-sector growth. Oxford Economic Papers, 65(1), 96–123.

Emani, C.K., Cullot, N., Nicolle, C. (2015) Understandable Big Data: A survey.

Computer Science Review, 17, 70-81.

Engel, T., Sadovskyi, O., Böhm, M., Heininger, R. and Krcmar, H. (2014) A Conceptual Approach for Optimizing Distribution Logistics using Big Data. Twentieth Americas Conference on Information Systems, pp. 1-9.

Fan, Y., Heilig, L. and Voss, S. (2015) Supply chain risk management in the era of big data. International conference of design, user experience, and usability (pp. 283-294). Springer International Publishing.

Foerstl, K., Reuter, C., Hartmann, E. and Blome, C. (2010) Managing supplier sustainability risks in a dynamically changing environment – Sustainable supplier management in the chemical industry. Journal of Purchasing & Supply Management, 16, 118-130.

Gencer, C. and Gürpinar, D. (2007) Analytic network process in supplier selection: A case study in an electronic firm. Applied Mathematical Modelling, 31, 2475-2486.

Gibbs, G.R. (2007) Analyzing qualitative data. SAGE Publications.

Guneri, A.F., Yucel, A. and Ayyildiz, G. (2009) An integrated fuzzy-lp approach for a supplier selection problem in supply chain management. Expert Systems with Applications, 36, 9223-9228.

Hallikas, J., Karvonen, I., Pulkkinen, U., Virolainen, V-M. and Tuominen, M. (2004) Risk management processes in supplier networks. International Journal of Economics, 90, 47-58.

Hashem, I.A.T., Chang, V., Anuar, N.B., Adewole, K., Yaqoob, I. and Gani, A. (2016). The role of big data in smart city. International Journal of Information Management, 36, 5, 748–758.

Ho, W., Xu, X. and Dey, P.K. (2010) Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. European Journal of Operational Research, 202, 16-24.

Huda, M., Maseleno, A., Atmotiyoso, P., Siregar, M., Ahmad, R., Jasmi, K.A. and Muhamad, N.H.N. (2018) Big data emerging technology: Insights into innovative environment for online learning resources. International Journal of Emerging Technologies in Learning (iJET), 13, 1, 23-36.

Joshi, K., and Yesha, Y. (2012) Workshop on analytics for big data generated by

healthcare and personalized medicine domain. In Proceedings of the 2012 conference of the center for advanced studies on collaborative research (pp. 267–269). IBM Corp.

Kahraman, C., Cebeci, U. and Ulukan, Z. (2003) Multi-criteria supplier selection using fuzzy AHP. Logistics Information Management, 16, 6, 382-394.

Kaisler, S., Armour, F., Espinosa, J. A., and Money, W. (2013). Big data: Issues and challenges moving forward. In 2013 46th Hawaii international conference on system sciences (HICSS), 7–10 January (pp. 995–1004). Wailea, Maui, HI: IEEE.

Kokangul, A. and Susuz, Z. (2009) Integrated analytical hierarch process and mathematical programming to supplier selection problem with quantity discount. Applied Mathematical Modelling, 33, 1417-1429.

Koskinen, I., Alasuutari, P. & Peltonen, T. (2005) Laadulliset menetelmät kauppatieteissä. Jyväskylä, Gummerus.

Lamba, K. and Singh, S.P. (2018) Dynamic supplier selection and lot-sizing problem considering carbon emissions in a big data environment. Technological Forecasting and Social Change.

Lester, A. (2014) Project Management, Planning, and Control: Managing Engineering, Construction and Manufacturing Projects to PMI, APM and BSI Standards. 6. p. Oxford, Butterworth-Heinemann.

Li, G-D., Yamaguchi, D. and Nagai, M. (2007) A grey-based decision-making approach to the supplier selection problem. Mathematical and Computer Modelling, 46, 573-581.

Liao, C-N. and Kao, H-P. (2010) Supplier selection model using Taguchi loss function, analytical hierarchy process and multi-choice goal programming. Computers & Industrial Engineering, 58, 571-577.

Manuj, I. and Mentzer, J.T. (2008) Global supply chain risk management. Journal of Business Logistics, 29, 1, 133-155.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A.H. (2011) Big Data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, pp. 1-137.

Mavi, R.K., Goh, M. and Mavi, N.K. (2016) Supplier selection with Shannon entropy and fuzzy TOPSIS in the context of supply chain risk management. Procedia – Social

and Behavioral Sciences, 235, 216-225.

Mayer-Schoenberger, V. and Cukier, K. (2013). “Big data: a revolution that will transform how we live, work, and think.” Eamon Dolan / Houghton Mifflin Harcourt.

Mensah, P., Merkuryev, Y., Klavins, E. and Manak, S. (2017) Supply chain risks analysis of a logging company: conceptual model. Procedia Computer Science, 104, 313-320.

Metsämuuronen, J. (2008) Laadullisen tutkimuksen perusteet. 3rd Edition. Jyväskylä, Gummerus.

Militaru, G., Pollifroni, M. & Ioanid, A. (2015) Big data in supply chain management: An exploratory story. Network Intelligence Studies, Volume 3, Issue 2 (6), 103-108.

Mitra, S., Karathanasopoulos, A., Sermpinis, G., Dunis, C. and Hood, J. (2015) Operational risk: Emerging markets, sectors and measurement. European Journal of Operational Research, 241, 122-132.

Mohtasham, K., Aziz, F.A. and Ariffin M.K.A.B.M. (2016) Vendor Selection Risk Management Framework in Automotive Industry. International Journal of Mechanical Engineering and Applications, 3, 3-1, 57-66.

Narasimhan, R. and Talluri, S. (2009) Perspectives on risk management in supply chains. Journal of Operations Management, 27, 2, 114–118.

Ng, Q.L. (2008) An efficient and simple model for multiple criteria supplier selection problem. European Journal of Operational Research, 186, 1059-1067.

Nyman, R., Gregory, D., Kapadia, S., Ormerod, P., Tuckett, D. and Smith, R. (2018) News and narratives in financial systems: Exploiting big data for systematic risk assessment.

Patra, K. and Mondal, S.K. (2015) Multi-item supplier selection model with fuzzy risk analysis studied by possibility and necessity constraints. Fuzzy Information and Engineering, 7, 451-474.

Patra, K. and Mondal, S.K. (2016) A supplier selection model with fuzzy risk analysis using the balanced solution technique with a soft set. Pacific Science Review A: Natural Science and Engineering, 18, 162-168.

Peat, M. (2013) Big data in finance. InFinance: The Magazine for Finsia Members, 127, (1), 34.

Peck, H. (2006) Reconciling supply chain vulnerability, risk and supply chain management. Int. J. Logist. Res. Appl. 9 (2), 127-142.

Philip Chen, C. and Zhang, C.Y. (2014) Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314–347.

Raguseo, E. (2018) Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. International Journal of Information Management, 38, 187-195.

Ranjan, R. (2014) Modeling and simulation in performance optimization of big data processing frameworks. IEEE Cloud Computing, 1, 4, 14-19.

Rao, C., Xiao, X., Goh, M., Zheng, J. and Wen, J. (2017) Compound mechanism design of supplier selection based on multi-attribute auction and risk management of supply chain. Computers & Industrial Engineering, 105, 63-75.

Richard, L., V., Matthew, E. and Carl, W., O. (2011). “Big data: What it is and why you should care.” IDC.

Riessman, C. K. (2008) Narrative methods for the human sciences. Sage.

Robson, C. (1993) Real World Research. Oxford: Blackwell.

Rozados, I.V. and Tjahjono, B. (2014) Big data analytics in supply chain management: Trends and related research. 6th International Conference on Operations and Supply Chain Management, Bali.

Ruhrmann, S., Hochdörffer, J. and Lanza, G. (2014) Assessment of dynamics and risks in supplier selection process. Proceria CIRP, 25, 2-9.

Salo, I. (2013) Big data – Tiedon vallankumous. Jyväskylä, Docendo Oy.

Salo, I. (2014) Big data & pilvipalvelut. Jyväskylä, Docendo Oy.

Sanayei, A., Mousavi, S.F. and Yazdankhah, A. (2010) Group decision making process for supplier selection with VIKOR under fuzzy environment. Expert Systems with Applications, 37, 24-30.

Schlegel, G.L. (2014) Utilizing big data and predictive analytics to manage supply chain risk. The Journal of Business Forecasting, 33, 4, 11-17.

Schonlau, M. (2002) The clustegram: A graph for visualizing hierarchical and non-hierarchical cluster analyses. The Stata Journal, 2, 4, 391-402.

Singh, A., Jain, D., Mehta, I., Mitra, J. and Agrawal, S. (2017) Application of big data in supply chain management. Materials Today: Proceedings, 4, 1106-1115.

Sivarajah, U., Kamal, M.M., Irani, Z. and Weerakkody, V. (2017) Critical analysis of big data challenges and analytical methods. Journal of Business Research, 70, 263-286.

Sonmez, M. (2006) A review and critique of supplier selection process and practices. Loughborough University.

Souza, G.C. (2014) Supply chain analytics. Business Horizons, 57, 595–605.

Svachko, K., Kuang, H. and Radia, S. (2010) The Hadoop Distributed File System. Mass Storage Systems and Technologies, 1-10.

Tiwari, S., Wee, H.M. and Daryanto, Y. (2018) Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Computers & Industrial Engineering, 115, 319-330.

Torres-Ruiz, A. and Ravindran, A.R. (2018) Multiple criteria framework for the sustainability risk assessment of a supplier portfolio. Journal of Cleaner Production, 172, 4478-4493.

Tracy, S.J. (2010) Qualitative quality: Eight big-tent criteria for excellent qualitative research. Qualitative Inquiry, 16, 10, 837–851.

Trkman, P. and McCormack, K. (2009) A conceptual model for managing supply chain risk in turbulent environment. International Journal of Production Economics, 119, 2, 247-258.

Tummala, R. and Schoenherr, T. (2011) Assessing and managing risks using the Supply Chain Risk Management Process (SCRMP). Supply Chain Management: An International Journal, 16, 6, 474-483.

Versace, M., and Karen, M. (2012) The case for big data in the financial services industry. IDC Financial Insights, 1–13.

Víegas, F.B., Wattenberg, M. and Dave, K. (2004) Studying cooperation and conflict

between authors with history flow visualizations. Proceedings of the SIGHI Conference on Human Factors in Computing Systems, 575-582.

Waller, M.A., and Fawcett, S.E. (2013) Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. Journal of Business Logistics (34:2), pp. 77-84.

Wang, Y., Kung, L. and Byrd, T. (2016) Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting & Social Change.

Wang, G., Gunasekaran, A., Ngai, E.W.T. and Papadopoulos, T. (2016) Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International Journal of Production Economics, 176, C, 98-110.

Wang, L. and Alexander, C.A. (2015) Big data driven supply chain management and business administration. American Journal of Economics and Business Administration, 7, 2, pp. 60.

Wu, K., Bethel, W., Gu, M., Leinweber, D., and Ruebel, O. (2013) A big data approach to analyzing market volatility. Algorithm Finance, 2(3–4), 241–267.

Wu, K-J, Liao, C-J, Tseng, M-L, Lim, M.K., Hu, J. and Tan, K. (2016) Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties, Journal of Cleaner Production.

Xiao, Z., Chen, W. and Li, L. (2012) An intergrated FCM and fuzzy soft set for supplier selection problem based on risk evaluation. Applied Mathematical Modelling, 36, 1444-1454.

Yang, Y.-C. (2011) Risk management of Taiwan’s maritime supply chain security, Safety Science, 49, 3, 382–393.

Yousafzai, Abdullah, et al. (2016) Multimedia augmented m-learning: Issues, trends and open challenges. International Journal of Information Management, 36, 5, 784–792.

Zeydan, M., Colpan, C. and Cobanoglu, C. (2011) A combined methodology for supplier selection and performance evaluation. Expert Systems with Applications, 38, 2741-2751.

Zhong, R. Y., Huang, G. Q., Lan, S. L., Dai, Q. Y., Xu, C., & Zhang, T. (2015). A big

data approach for logistics trajectory discovery from RFID-enabled production data. International Journal of Production Economics, 165, 260–272.

Zhong, R.Y., Newman, S.T., Huang, G.Q. and Lan, S. (2016) Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. Computers & Industrial Engineering.

Zsidisin, G.A. (2004) A grounded definition of supply risk. Journal of Purchasing & Supply Management, 9, 217-224.

Websites

Corea, F. (2016) Big data & risk management in financial markers (Part I). Medium. [online document]. [Accessed 6.6.2018] Available: https://medium.com/@Francesco_AI/big-data-and-risk-management-in-financial-markets-part-i-eed4e245f3db

Curtis, P. (2016) Risk modeling: Finding the right approach. The Wall Street Journal. [online document]. [Accessed 4.3.2018] Available: https://deloitte.wsj.com/riskandcompliance/2016/08/16/risk-modeling-finding-the-right-approach-3/

GEP (2018) Big data analytics for procurement. [online document]. [Accessed 18.8.2018] Available: https://www.gep.com/big-data-analytics-procurement

McDaniel, W. (2016) Risk-adjusted supply chains help companies prepare for the inevitable. [online document]. [Accessed 4.3.2018] Available: http://demand-planning.com/2016/02/19/risk-adjusted-supply-chains-help-companies-prepare-for-the-inevitable/

Meraglim (2017) The importance of big data analysis for risk management. [online document]. [Accessed 20.9.2018] Available: https://www.meraglim.com/blog/financial-data-analytics/the-importance-of-big-data-analysis-for-risk-management/

Parasuraman, K. and Michel, A. (2017) Big Data in Catastrophe and Supply Chain Risk Modeling. Supply Chain Navigator, January 2017. [online document]. [Accessed 4.3.2018] Available: http://scnavigator.avnet.com/wp-content/uploads/2017/01/Big-Data-in-Catastrophe-and-Supply-Chain-Risk-Modeling.pdf

Swaminathan, S. (2012) The effects of big data on the logistics industry: PROFIT

ORACLE Technology Powered. Business Driven Available http://www.oracle.com/us/corporate/profit/archives/opinion/021512-sswaminathan-1523937.html.

APPENDICES

Appendix 1. Interview questions

Interviewee's information:

- Name
- Title

Supplier selection process:

1. How critical and important decision is the supplier selection to you?
2. Do you use some kinds of specific methods or models for supplier selection?
3. What are your supplier selection process steps?
4. What are your supplier selection criteria and measurements you use? Which ones are the most important ones for you?
5. How do you get or gather information on your suppliers to make the selection decision?
6. Who does the final decision about supplier selection? (Who has the responsibility?)
7. What kind of challenges do you encounter with the supplier selection process?

Risk management in supplier selection:

8. What supplier risk evaluation and monitoring methodologies or models do you have in place at the moment?
9. Have you some kind of supplier risk management process (steps) in place? How do you strive to prevent risks from realizing?
10. Who is responsible for risk management in your company?
11. What kind of risks do you encounter or have encountered with your suppliers or business operations previously?
12. What do you think are your greatest supplier risks you encounter?
13. Do you have any challenges and problems related to risk management?

Big data:

14. How would you determine the term big data?

15. You do not utilize big data in your business that much, why is that? Are there some obstacles for the usage or why haven't you seen big data usage that necessary?
16. Do you see big data usage as an opportunity or rather as a challenge? Why?
17. What are/could be your most important big data channels or sources for you?
18. Do you have some big data technologies in use, what?

Integrating big data to risk review related to supplier selection:

19. Do you share information and data within your supply chain or would you see this as a possibility?
20. How dependent are on your supplier performance and do you have close relationships with them?
21. What challenges do you see in using big data for supplier selection and risk management related to the selection decision?
22. Do you think you have the requirements that big data usage requires? The knowledge, the skills, right technologies and open attitude?
23. How would possible big data usage change your operations and organization? What changes do you think you would have to make to your organization and to your operations to be able to utilize big data in risk management and supplier selection processes? Would it require new resources, training etc.?
24. Where in risk management and supplier selection would you see a possibility to use big data? In which process phases?