Aalto University
School of Science
Degree Programme in Industrial Engineering and Management

Amir Mostafa Heidari

Exploration of Big Data in Procurement - Benefits and Challenges

Master's Thesis

Espoo, April 23, 2018

Supervisor: Professor Kari Tanskanen

Thesis advisor: Doctor Riikka Kaipia

Author Amir Mostafa Heidari

Title of thesis Exploration of Big Data in Procurement – Benefits and Challenges

Master's programme Industrial Engineering and Management

Thesis Supervisor Professor Kari Tanskanen

Major or Minor/Code Operations and Service Management / SCI 3049

Department Industrial Engineering and Management

Thesis Advisor D.Sc. Riikka Kaipia

Date 23.04.2018 **Number of pages** 67 **Language** English

Abstract

Emergence of Big Data had positive implications in various industries and businesses. Big Data analytics provides the ability to harness massive amount of data for decision making purposes. One of the important use case of Big Data analytics is in supply chain management. Increased visibility, enhanced bargaining position in negotiations, better risk management and informed decision making are examples of benefits gained from Big Data analytics in supply chain. Although there are advances in analytics application throughout supply chain management, sourcing applications are lagging behind other functions of supply chain. The purpose of this study is to analyse use cases of exploiting Big Data for purchasing and supply purposes, in order to help companies having more visibility over the supply market. Data collection in this study was carried out through the use of semi-structured interviews which then were coded and categorized for comparison.

The results pointed out that big data aids in identifying new suppliers. Additionally, having transparency over n-tier suppliers for managing risks were important for companies. Most of the companies are using descriptive analytics. However, they expected to have predictive analytics to become aware of market situation and gain better position in negotiations. Furthermore, this research showed that to prevent supply disruptions, the Big Data analytics should send timely warnings to managers. The main expectations from Big Data analytics are gaining transparency, automation of data collection and analysis, prediction, availability of new data sources, more efficient KPIs and better representation of data. The main hurdle in Big Data initiative is unintegrated and non-homogenous internal data.

Keywords Big Data analytics, supply management, procurement, supply chain, sourcing

Acknowledgement

I wish to thank my supervisor professor Kari Tanskanen for introducing me to this project and for having his office door open whenever I had a question. Also I would like to thank my advisor doctor Riikka Kaipia for her guidance throughout the process of this thesis and for her aid in finding case companies of this research. Finally I would like to thank my family for their unconditional support throughout my years of study.

Table of Contents

List of figures and tables

Figures

Tables

List of abbreviations

AI	Artificial Intelligence
B2B	Business to Business
BI	Business Intelligence
BI&A	Business Intelligence Analytics
BOM	Bill of Material
CPFR	Collaborative Planning, Forecasting and Replenishment
CPO	Chief Procurement Officer
CRM	Customer Relationship Management
DPB	Data Science, Predictive Analytics and Big Data
EDI	Electronic Data Interchange
ERP	Enterprise Resource Planning
HP	Hewlett-Packard
HR	Human Resources
IBM	International Business Machines Corporation
IT	Information Technology
KPI	Key Performance Indicator
OLAP	Online Analytical Processing
PO	Purchase Order
RFx	Request for Information/Proposal/Quotation
ROI	Return on Investment
SCM	Supply Chain Management
SVP	Senior Vice President
TCO	Total Cost of Ownership

1 Introduction

Big Data is the hot trend in today's world that can benefit many industries and businesses. In this study the focus has been particularly set on potential benefits that procurement and sourcing can reap from Big Data analytics implementation but first we need to understand what Big Data is. There are many definitions for Big Data. The Big Data term comes from massive volume of accessible data, that has high velocity in generation of data and transferring that comes from variety of different sources. These sources of data range from omnipresent smart devices of consumers to manufacturing and logistics processes that can be captured thanks to IoT technology. Although Big Data is no longer considered a black box among firms, still its usefulness is not fully understood. Big Data is being exploited for scientific and industry related purposes and supply chain management is no exception. Considering supply chain from upstream players to downstream players, there are lots of operations that generate data which can be used for improving efficiency, gaining transparency and finally leading to competitive advantage. Walmart and Amazon are examples of firms with modern supply chains that implemented analytics capability in their supply chains. Nevertheless, not every firm has adopted this technology. In addition to that, not all functions of supply chain make use of Big Data with same extent.

Companies are focusing more on their core competencies and outsource their non-core operations to gain better margins. With globalization, outsourcing is going beyond continents. Design can be done in one part of world, manufacturing in another low cost country and distribution among other countries in the world. Considering that substantial percentage of a firm's resources are external resources, its importance has moved procurement from a function working in silo to a strategic function. Procurement is becoming a central function that needs to be coordinated with all other functions. Furthermore, significant amount of a firm's revenue goes for its procurement. This exhibits importance of procurement in a firm's bottom-line. Therefore, there are great opportunities for procurement to make use of data. Procurement is behind other functions of a firm in terms of analytics and recently it is gaining more attention.

This research tries to find out how Big Data capabilities can help procurement in sounder decision making. One of the studied aspects was supply base, in order to understand if there are opportunities in identification of new suppliers and monitoring more than immediate suppliers in value chain. Managing suppliers in a global supply chain is a challenging task that Big Data can provide transparency and control towers to facilitate it. Another studied aspect is

having access to greater pool of data and what kind of potential data procurement can benefit from. Supply professionals were interviewed to understand what they expect from their procurement systems and how Big Data can contribute to that. What kind of transparency is needed? What needs to be automated? What representation of data is useful? Furthermore, how Big Data can help in procurement risk management. Additionally, the value of social media and issue of trusting data analytics were studied. Nevertheless, there are hurdles in implementation of Big Data analytics that required consideration.

1.2 Research gap

Researches conducted on Big Data's impact on supply chain have mostly addressed the issue with a holistic view of supply chain. With presence of smart phones and IoT sensors, great focus has been placed on modelling behaviour of customers on demand side, logistics and manufacturing efficiency. When it comes to Big Data and advanced analytics, procurement as mentioned by Sanders (2016) is behind development and research compared to other functions of supply chain even though substantial amount of revenue goes for procurement. Certainly, for a company to succeed in its analytic journey, all of its functions should keep up with each other in implementing Big Data analytics. The literature that addresses Big Data analytics in procurement are quite limited. Sanders (2016) explains Big Data benefits and hurdles in the whole supply chain and all functions of a firm. However, he touches procurement briefly. Richey et al (2016) explores Big Data in supply chain through case studies and studies procurement briefly. Nevertheless, the success factors of analytics identified by Richey et al (2016) is true in Big Data analytics of procurement as well. Similarly, Weng et al (2016) studies Big Data in logistics and supply chain with a brief note on procurement. (Sanders, 2014, p.131-147) has a chapter related to Big Data in sourcing which he identifies some use cases of Big Data in procurement. In terms of risk management Fan et al (2015) also studies risk management in context of supply chain.

The literature related to Big Data in procurement is mostly part of a bigger study in supply chain. The main gap in literature is lack of case studies specifically dedicated to Big Data in procurement. Consequently, this master's thesis was conducted to get in touch with purchasing professionals through case study and tries to see if similar findings or new findings emerge.

1.3 Research objective and research questions

This purpose of this study is to analyse use cases of exploiting Big Data for purchasing and supply purposes, in order to help companies in having more visibility over the supply market. Furthermore, we are looking to elicit needs of supply managers regarding what they expect to gain from intelligent use of data and what intelligence they need to make sound strategic decisions that improves procurement management and lowers supply risk. In order to reach this objective following main research questions was created:

> **Research Question 1:** *What potential uses and needs companies are having in the area of procurement that can be supported with Big Data and advanced analytics?*
>
> **Research Question 2:** *What are the challenges and hurdles in implementing Big Data analytics in procurement?*

1.4 Research methodology

Case study method was chosen for this research. As mentioned earlier in research gap since this topic has not been investigated sufficiently earlier, an explorative qualitative approach was adopted to discover new insights about the topic. Case study provides the possibility to look at the phenomenon from various viewpoints and the qualitative nature of study removes constraints of rigid data collection. This research follows the abduction approach since it has a combination of inductive and deductive approach. Initially literature related to topic was identified and reviewed and results were deducted based on empirical data. Empirical data collection was carried out through multiple case studies in form of semi-structured interviews of purchasing and supply management professionals of case companies. Data was coded, analyzed and discussion and conclusion was written based on it.

1.4 Structure of the study

This study is divided into five chapters. In chapter 1, the topic is introduced and objective and importance of study is highlighted. Chapter 2 reviews literature. It starts by defining Big Data analytics followed by its role in supply chain management and finally its role in procurement. Chapter 3 describes research methodology, data collection and data analysis process. Chapter

4 presents the results of the study. Chapter 5 draws conclusions and implications while pointing out limitations of study and potential for further research.

2 Literature review

2.1 What is Big Data?

Big data is the hot trend in today's world and companies are trying to harness this concept to leverage it for their business needs. However, companies do not fully understand what actually the Big Data is. It seems that the solution to better performing businesses rests in Big Data, but it is not clear why and how. In order to exploit Big Data to its full potential we need to begin with a thorough understanding of this concept. Forbes's author, Press (2014) suggests some definitions for big data. Two of its relevant definitions are "The shift (for enterprises) from processing internal data to mining external data" and "A new attitude by businesses, non-profits, government agencies, and individuals that combining data from multiple sources could lead to better decisions". So when it comes to Big Data, both internal and external data are important and companies are starting to see value in analysing these data. Manyika (2011) sees importance of Big Data in competition, firm's growth and innovation.

Data is generated in our everyday activities. It is no longer limited to mobile networks or communication and spreadsheets. With digitalization and integration of technology in our day to day life, massive amount of data are collected and this is can be reason for the term "Big Data". From crowd-funded business models and gadgets, large scale banking and mobility to individuals' data and our behavioural patterns can be recorded and analysed for improvement of performance or preventing delays and even threats (Koutroumpis & Leiponen, 2013).

Gartner analyst Doug Laney coined 3 attributes to big data: volume, velocity and variety (Laney, 2001). Nevertheless, Big Data attributes have been evolved. Vossen (2014) considers 4Vs for Big Data and Demchecnko et al (2013) proposes 2 additional Vs to the main 3Vs: value and veracity.

Volume

Clearly the first thing that justifies volume attribute, is the term "Big Data" itself. Volume is the most important feature of big data. This is due to the fact that it is possible to collect data from wide range of sources. These sources are no longer limited to internal data. Data can be collected from sources such as transactions, sensors, behaviours and social networks. Possibility to collect data from various sources contributes to increasing volume of data. Data needs to be "be accessible, searchable, processed and manageable" (Demchenko et al, 2013).

As an example, social media like Facebook produce huge amount of data on a daily basis that can be exploited in global scale. What is considered Big Data might not be the same in future because of increase in capacity and data generation, also the type of data and industry generating the data makes it difficult for defining magnitude of volume (Gandomi et al, 2015).

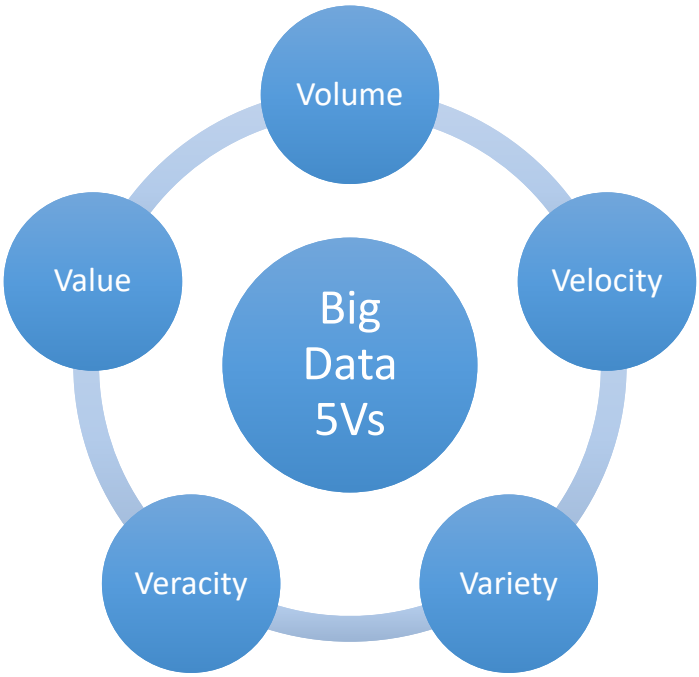

Figure 1 Big Data 5Vs (modified from Demchenko et al, 2013)

Velocity

Velocity refers to speed of generation of data and analysis of it. With prevalence of smart devices and sensors, huge amount of data are being generated that need real-time analytics. Living in time of internet of things, massive amount of data cannot be managed by traditional data management tools and requires big data technologies for storing and analysis (Gandomi et al, 2015).

Variety

Variety describes heterogeneity of data. With technology advances, firms can collect both structured and unstructured data. Structured data in form of spreadsheets and relational databases only comprise 5% of data. Unstructured data are in form of text, audio, image and video (Gandomi et al, 2015). Additionally, collecting data from biological, human and societal systems increases complexity. Furthermore, introducing data technologies in non-digital fields such as psychology and behaviour research leads to further unstructured data (Demchenko et al, 2013).

Veracity

Veracity was coined by International Business Machines Corporation (IBM) for reliability of data. For example customer emotions in social media can be biased or judgemental. On the other hand it can contain valuable insights (Gandomi et al, 2015). Veracity has two main components: consistency defined by statistical reliability and trustworthiness that is determined by factors such as origin, reputation and reliability of processing method. Veracity ensures trustworthiness of data in its whole life cycle from collection to processing and storage. Factors that need to be taken into account are integrity, authenticity, identification of sources, timeliness and accountability (Demchenko et al, 2013)

Value

Value dimension of Big Data was introduced by Oracle. Data has low relative value to its volume (Gandomi et al, 2015). Added value is the value that is gained through processing of data. It can bring benefit to processes or can be in form prediction or forecast (Demchenko et al, 2013). Manyika (2011) presents a broader range of possibilities to add value rather than merely prediction and forecast. Big Data provides transparency of information and more frequent usage. Accurate data collection capability can boost performance and enable performing experiments and doing forecasts. Furthermore, customer segmentation will be more detailed for more tailor made products. In the end, analytics capabilities lead to better decision making.

In addition to these main 5Vs, there are variability and complexity. "Variability refers to the variation in the data flow rates" (Gandomi et al, 2015). Complexity arises from existence of many sources of data which makes cleansing it necessary. So far it seems clear that there is no absolute definition for big data and the further we go ahead the more dimensions are seen and maybe there will be more Vs in future as the need for them is discovered.

2.2 Business analytics and Big Data analytics

Holsapple et al (2014) tries to define business analytics in form of a framework with some building blocks. The more technical block blocks of this framework are not discussed since they are not in scope of this thesis. The first building block is **movement**. In any entity such as an organization or supply chain, the movement is similar to working culture that welcomes

usage of analytics. It is grounded in identifying and solving problems through evidence. Simply evidence is not considered as an add-on next to politics that can be overlooked. So the movement is supposed to build a culture that considers evidence as the driver of decision making. This culture fosters creativity, collaboration and realistic view of situations. However, building such philosophy and culture around analytics with a plan on sustaining it and defining key performance indicators (KPI), is a challenge of its own.

Another building block is **collection of practices and technologies**. This part can exist without the previous block, the movement but it might not be effective enough. In this part lies techniques and technologies that are used for evidence based operation to derive knowledge for decision making. Business analytics is usually viewed by this part, techniques and practices Next building block is **transformational process**. At this stage the practices and techniques turn evidence into understanding and action. Business analytics has a **set of capabilities**. These capabilities are usage of statistical and techniques and working with descriptive, predictive and prescriptive analytics that provides evidence for decision making. These types of analytics will be discussed more extensively further ahead. There are examples of putting these capabilities to work: interpretation of competitors' actions, distilling message in big data, optimization, identifying suppliers' strategies and new sources of data (Holsapple et al, 2014).

The final important building block is **decisional paradigm**. Business analytics is considered as an approach in decision making alongside other approaches. These approaches can be in line with each other or can be conflicting. The issue that is raised here is assessing whether this decisional paradigm is appropriate for an organization considering analytics movement and presence of required capabilities. If it is appropriate then design of transformational processes can proceed. Also it is important to think about aligning business analytics with other decisional paradigms and avoiding its misuse. Holsapple et al (2014) explains that business analytics try to achieve competitive advantage and support organization's strategic goals by informed decision making that yields better results. Nevertheless, it should be noted that owning Big Data analytics on its own does not bring competitive advantage to a firm. It is merely a technology that facilitates decision making and its performance is largely dependent on quality of data and algorithms in use. In order to move in direction of gaining competitive advantage informed decision making should be accelerated. Therefore, it is necessary to gain buy in for Big Data analytics from managers in form of cultural movement mentioned by Holsapple et al (2014). Additionally, managers should be provided with guidelines to implement this technology in their businesses (Chan et al, 2014)

Considering data analytics, one can think of it as statistical analysis in form of spreadsheets or some automation through Excel. However when we are talking about Big Data, there are big databases that cannot be analysed manually. Big data analytics deals with massive data by technologies that have eased statistical analysis, data mining and machine learning algorithms. Clustering, classification and regression analysis can be done in large scale. Furthermore, optimization algorithms are easily applicable, thanks to these technologies (Chen et al, 2012).

2.3 Types of analytics

Chen et al (2012) divides business intelligence analytics (BI&A) to 3 levels: BI&A 1.0, BI&A 2.0 and BI&A 3.0.

BI&A 1.0 relies on structured data that is extracted and collected through legacy systems present in companies and analysis is based on statistical analysis and data mining techniques. Sallam et al (2011) considers following capabilities for BI&A 1.0: reporting, dashboards, *ad hoc* query, search-based Business intelligence (BI), online analytical processing (OLAP), interactive visualization, scorecards, predictive modelling, and data mining.

BI&A 2.0 is result of emergence of search engines like Google and Yahoo which enabled online presentation of businesses and their interaction with other businesses and customers. Web 2.0 services has helped in collection and generation of user contents in social media such as forums and social networks. This has created opportunity for bidirectional communication between customer and business. Some technologies used in this area are: text mining, web mining and social network analysis. BI&A 3.0 is related to prevalence of smart phones and devices. With internet of things, sensors and RFIDs, there are great opportunities for analysis of large scale mobile data.

Analytics are divided into 3 categories: descriptive, predictive and prescriptive.

Descriptive analytics is used whenever necessary. It tries to identify problems and opportunities in the existing processes and functions by techniques such as OLAP and drill down (Wang et al, 2016). It uses historical data to identify patterns and modelling past behaviour (Assunção et al, 2015). Questions that are answered are what has happened in past? What trends existed and what were the exceptions? After obtaining insights from data, it will give alerts and it is capable of visualizing results in a format that is easy to understand (Rozados et al, 2014).

"**Predictive analytics** attempts to predict the future by analysing current and historical data" (Assunção et al, 2015). It has the technology to learn from data thanks to machine learning and data mining algorithms. Time series analysis, advanced forecasting of safety stocks, regression and statistical analysis are examples of techniques used in this type of analytics (Rozados et al, 2014). Waller et al (2013) defines predictive analytics in supply chain as:

Supply chain management (SCM) "predictive analytics use both quantitative and qualitative methods to improve supply chain design and competitiveness by estimating past and future levels of integration of business processes among functions or companies, as well as the associated costs and service levels".

Prescriptive analytics use the predictions in order to propose actions in situations that will be faced in future whether to benefit from them or avoid risks. Optimizations, simulations, what/if scenario analysis and similar techniques that serve the purpose of dealing with uncertainties is related to this domain (Rozados et al, 2014). This type of analytics is the least common among firms. Additionally, trusting a system with making decision is still subject of consideration in minds of practitioners.

2.4 Big Data analytics in supply chain management

The aforementioned capabilities of big data can be leveraged in supply chain management. From upstream to downstream of supply chains, massive data is generated in each operation and supplier in the chain. Provision of accurate and real time insights help supply chain professionals to continually identify and respond to supply chain issues. These data are in form of structured, semi structured and unstructured data (Rozados et al, 2014).

Figure 2 shows taxonomy of data in supply chain. It is evident that structured data is mostly present in internal systems and has lower volume and velocity. As we go further in horizontal axis variety increases as data becomes more unstructured. At the same time volume and velocity surge. Even though enterprise resource planning (ERP) and customer relationship management (CRM) data are core data, they only compromise a small fraction of whole data sources available to use.

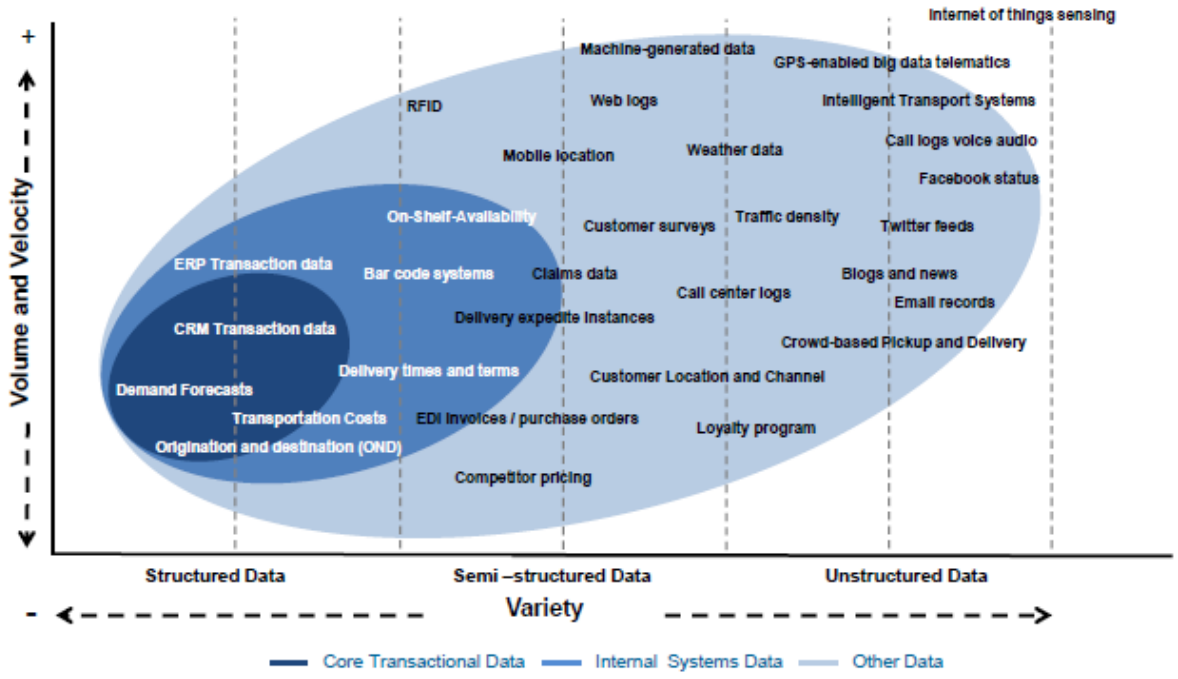

Figure 2 Taxonomy of data in supply chain (Rozados et al, 2014)

Gunasekaran et al (2017) argues that big data analytics can aid supply chain through improving visibility, resilience and robustness. It generates cost saving and contributes to competitiveness of a firm. Additionally, firms can use it to beat competition by increased transparency, effective decision making and customer segmentation provided by data capabilities.

2.4.1 Implications of 3Vs in supply chain

Hofmann (2017) argues that out of the 3Vs of big data, velocity plays the biggest role in diminishing bullwhip effects. The data should be captured, processed and transferred as fast as possible not only in a single firm but among all members of supply chain. Due to importance of speed and agility in decision making, velocity enables increased effectiveness in decision making. In a long supply chain that number of intermediaries increases, velocity benefits last upstream player. Richey et al (2016) conducted a study inquiring about Big Data from supply chain professionals. He mentions that most of the respondents noted the increased velocity of the data being produced. Supply chain managers considered velocity of data as both an obstacle and an opportunity.

As mentioned earlier, emergence of smart devices, sensors, RFIDs and internet of things has led to massive generation of data of different sort. Data comes in different formats and different level of structure. So the challenge that variety imposes is, finding new ways of storing and

processing data since old databases are only suitable for one type of data. The challenge of volume is mostly evident in multi staged supply chains that there is the possibility of missing volume since data is running across multiple platforms (Hofmann, 2017).

Traditional applications prior to big data only dealt with transactional data and could not react fast to unexpected supply chain events. Therefore, big data applications are needed to make use of these massive data to expedite decision making and bestow agility to supply chain (Hofmann, 2017).

2.4.2 Key success factors of implementing Big Data in supply chain

In Richey et al (2016) supply chain professionals are asked to determine what success factors of implementing Big Data in supply chain management are. The first factor is **SC systems integration.** They expected improvement in supply chain through an integrated system that they can communicate with suppliers in the same system and leveraging big data capabilities to link that system to market needs. Schoenherr & Speier-Pero (2015) also identifies lack of integration as a barrier to Big Data. Nevertheless, an integrated system incurs cost. The two main costs related to that are buying equipment and training. Although such system can reduce waste, presence of diverse system in supply chain creates data redundancy waste and having fewer systems helps in quality of information. Additionally, it is much easier to build systems from ground zero rather than having multiple systems in place, since connecting systems from different functions and with different format is a challenge of its own. Also Wang et al (2015) argues that great value of Big Data comes from integrated data sources.

Next factor on the list is **improved forecasting/decision making.** Traditional decision making is mostly based on relationships with suppliers and their background or track record. Big data provides the ability of decision making based on evidence as mentioned earlier. The important point about relationships is mentality. People are used to decision making based on experience and opinion of people they trust. Quantitative analysis is a new practice against this mentality (Richey et al, 2016). Reliance on relationships and experiences for decision making, displays the need for change management for buy in of new morality.

Even with Big Data, communication is key in sound decision making since decisions are intersection of human capital with advanced analytics capabilities. Experts need to interpret data and disseminate it to key stakeholders and disseminating it in a format which is easily

understandable. Forecasting provides opportunity for stronger informed decisions and ability to identify more risks (Richey et al, 2016).

Next success factor is **human capital**. Firms adopting Big Data considered management's mentality as hindrance in implementation because their emphasis is on relationships rather than quantitative data. These managers have usually a long tenure in the company and they are not used to data driven business. This is due to unavailability Big Data capabilities earlier in their career. In Schoenherr & Speier-Pero (2015), one third of respondents were not familiar with analytics which shows a need for educating people. Nonetheless, at lower level people are needed to put actions into practice. Therefore, people are needed who are knowledgeable in Big Data. Such people can also help in data collection and elevating data quality (Richey et al, 2016). Gao et al (2015) had also conducted a study on success factors of Big Data implementation. One dimension of their results was dedicated to role of people. They believed that current staff should have Big Data skills or people with such skills should be hired externally. Additionally, a multidisciplinary team from different department should take part in Big Data project. Although people who are trained in analytics are necessary for Big Data initiatives, business people who welcome data driven business should be present to reap benefits from available analyst resources.

Security and governance is a big concern for supply chain managers in terms of data ownership, data storage and data privileges. In spite of security risks many managers believe that sharing information provide benefits to all members of supply chain. Even if one member shares information, benefits are more than costs. However, not all information should be disclosed. A good example is pricing information and profit margin.

On the other hand not all the information is kept confidential for financial reasons. There are legal regulations for protecting data privacy and they are industry specific. Healthcare is one of these industries. Patients' information must remain confidential. Additionally data must be stored safely and this makes ensuring security of cloud storages and internal systems. Even though managers become willing of sharing information, legal requirements might be stricter. In terms of data ownership, many are concerned about who owns the data but in emerging economies data sharing is even stricter (Richey et al, 2016).

Given all the security risks and not a clear stance on what can be shared and what cannot be shared, many managers believe that data sharing decreases risks of decision making not only in top management but in entire organization. Big data is a "**potential tool for improving risk**

management efforts which presently are difficult to manage thoroughly". There are significant amount of risks associated with suppliers and big data helps in managing them (Richey et al, 2016).

Another challenge for supply chain managers is **storage** of massive data. While volume of data gives the potential of more analysis, not all of the data is useful. One thing to consider is short term and long term storage of data. Because it is not clear which data is useful in future. It cannot be determined what should be kept and what should not. So the challenge of physical storage is present when we are not aware what should be kept and if we delete data just for storage limitation, it might be of use in future. Nonetheless, there are costs related to infrastructure, protection and capacity. Industries that have poor system integration exacerbate this situation because more effort is needed for organizing data. While managers see opportunities in big data but still they do not exactly know what to do with it. Some managers believe as they start storing and working with data they learn what is useful and what is not but finally the know how to make use of data for decision making is important (Richey et al, 2016).

Big Data helps in gaining **operational efficiency.** The information gathered can be leveraged to improve all the players that are connected to data driven system. Data gathered not only improves supply chain partners but it is useful for individual functions within these partners e.g. procurement function. However, balancing costs of Big Data system and its potential benefits are challenging (Richey et al, 2016).

The final significant success factor is **partner transparency**. Supply chain partners want transparency for communication. The term "control tower" is relevant here. What they want is transparency of all plans and platforms and communicating with each other. Respondents believed that transparency is beneficial in gaining trust of supply chain partners since they rely on shared data with clear sources. (Richey et al, 2016).

Richey et al (2016) points out other factors that are were less important in view of their respondents but they are worth mentioning. Importance of data quality and having inputs from different people makes having **one version of truth** a necessity. This is can be solved by having an integrated system which is not in place in many cases. To have same interpretation from both operational people and strategic people, one version of truth is needed. Managers aspire to have simplified Big Data solutions that is accessible from any location at any time, however some of them believe that this capability should be developed internally and others believe in third party providers. Finally, to achieve one version of truth databases must be cleansed and

quality must be improved. As mentioned earlier massive data is available but not every firm knows how to make use of it. They either lack the technology or lack data scientists. In addition to that distinguishing between useful and useless data is a challenge. Therefore, firms have to **climb the learning curve**. The goal of a business is optimizing its profits. Big Data and adopting **customer orientation** helps in this regard. It enables capturing customer views, predicting their needs and customizing products.

2.4.3 Benefits of Big Data in supply chain management

Wang et al (2015) has compiled lists of benefits of Big Data in supply chain management. Benefits are divided in 4 groups: visibility and accuracy of information, operational efficiencies, higher service quality and new business models and better prediction. Regarding visibility it enables product and service visibility, identification of problematic suppliers and problems for suppliers, giving early warnings and reduction of inventory and supply chain risk. Regarding operational efficiencies, it aids in expediting decision making, real time vendor management and real time view of demand and sale for expediting sourcing process. Service quality is improved through close interactions with customer. New business models lead to emergence of companies who are information driven and act as intermediaries. Schoenherr & Speier-Pero (2015) conducted a large scale survey of supply professionals to understand benefits of predictive analytics in SCM. Findings of Schoenherr & Speier-Pero (2015) confirm the list provided by Weng et al (2015) except that it does not mention emergence of business models. Additionally some more benefits are included. For example, greater power in relationship with suppliers and customers, agility in response to changing environment and better bargaining position in negotiations.

2.4.4 Challenges of Big Data in supply chain management

Regardless of general challenges of Big Data that are data management capabilities, privacy, data ownership and deriving meaningful insights from large volume of data, implementing it in each domain has its own unique challenges and requires domain knowledge. Therefore, a supply chain has its own challenges. Usually supply chains are not owned by one specific company, there are multiple number of players, each with their own fragmented system which requires standardization efforts. Many of stakeholders do not understand the value that can be gained by Big Data. Even though sharing information creates value, firms can be hesitant in this regard especially if they are concerned about losing competitive advantage. Finally shared

data requires a repository accessible for all partners and such repository currently do not exist (Wang et al, 2015).

Sanders (2016) sees four hurdles in implementation of Big Data analytics in supply chain management. The first hurdle is **needle in a haystack.** Companies are under influence of Big Data hype and are rushing to implement it not to fall behind the competition, however they are doing it randomly. Therefore relationships and causations found just through number crunching and without focus can yield false results.

The next hurdle is **islands of excellence**. Firms buy application for specific functions. These algorithms are efficient in optimizing the selected function, however if it is not in coordination with the whole supply chain, these efforts become futile. Leading supply chains such as Walmart's and Zara's have understood that optimization of functions in silo is insufficient. In their supply chain any information captured through sell side will be conveyed to all supply chain functions. So any shortage or delay will be coordinated with all of the functions. These advanced companies view supply chain as a system with interconnected functions. Optimization of one function cannot be done without consideration of other functions, since it can increase costs in other functions (Sanders, 2016).

Many companies are struggling with their measurement metrics. Sanders (2016) call it **measurement minutiae**. Companies have too many metric to cover everything but only a few know which one is important to focus on. Consequently, companies are moving towards creating fewer customized metrics that are easier to manage.

At last there is the issue of **analysis paralysis**. Companies have data sources and are generating data through supply chains. They are awash in data, however there is confusion of what to do with this data and where to start that has led to paralysis. They cannot make use of data even though new analytics technology are offered to them (Sanders, 2016). It implies that advanced analytics and technology is not the issue but the data itself is the issue and this issue ranges from structure and format of data to determining its relevancy.

Data quality is issue that should not be overlooked. Hazen et al (2014) mentions that data usefulness is dependent on quality since it will affect decision making. Costs of poor quality data in decision making can be from 8% to 12% of revenue for typical organization and up to 40% to 60% for service organizations. Furthermore, it causes job dissatisfaction and deteriorates trust in organization. There are four dimensions for quality of data: accuracy (error free), timeliness (up to date), consistency (same format) and completeness. "**The lack of**

common data format standards and the transfer of data between dissimilar systems" also undermines reliance of firms on data science, predictive analytics and Big Data (DPB) for growth and innovation (Hazen et al, 2014). Therefore, supply chain managers should be concerned about quality of data as much as they are concerned about their products quality.

2.4.5 Analytics application in supply chain

Amazon and Walmart are examples of companies that transformed their supply chains using Big Data. Companies such as Facebook, Google and LinkedIn can also use Big Data capabilities. But for others adopting this technology is costly and the return on investment (ROI) is uncertain. Additionally, firms have massive data but do not know what to do with it and where to start. Furthermore, Big Data analytics require competencies that might not be present in company so it has to be outsourced. With machine learning capabilities we no longer test human hypothesis but we test hypothesis that have not been generated by human (Sanders, 2016). Figure 3 exhibits application of analytics across supply chain.

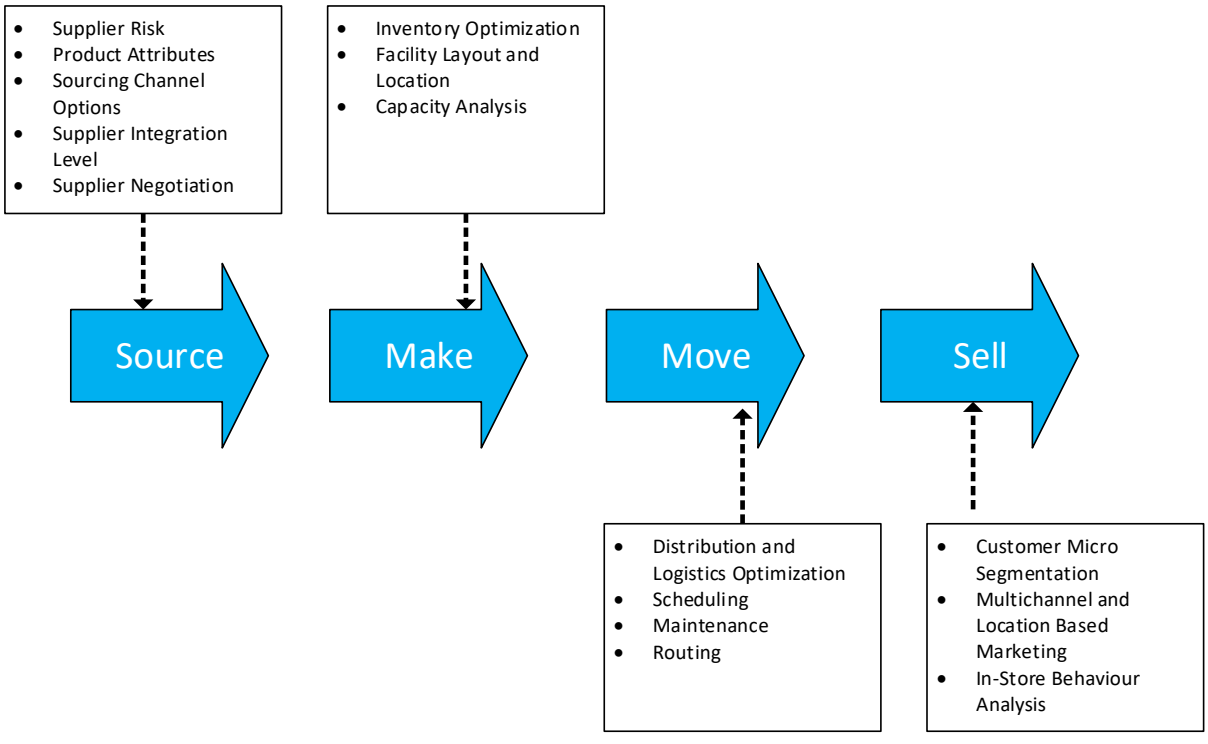

Figure 3 Analytics Across Supply Chain (modified from Sanders, 2016)

In this figure supply chain is divided in 4 categories: source, make, move and sell with areas that analytics can help. Each of these blocks has its own subgroups that discussing them is not in scope of this thesis. The concentration of this thesis is on sourcing.

2.5 Procurement and Big Data

Although Big Data is exploited in marketing, logistics and other areas of supply chain, focus of this part of literature review is on sourcing and procurement as this thesis is mostly concerned with Big Data's impacts on procurement. Since sourcing expenditure comprises of 50 to 90 % of a firm's revenue, there is great opportunity for leveraging Big Data to benefit this part of supply chain. Nevertheless, compared to logistics and operations, sourcing applications are lagging behind and it is expected to gain more attention from now on. Currently sourcing applications are used for segmenting suppliers based on key capabilities, identifying risks, cost measurement and managing tail-spend. Another use case is identifying buying behaviour of customers and using that for supplier negotiations (Sanders, 2016).

In the study conducted by Richey et al (2016), one supply chain professional believed that the advantage that Big Data can provide to procurement is **providing trends of procurement**. Based on these trends consolidation and pulling of the prevailing volume would give **strength in bargaining, in negotiation, improvement of quality, better delivery and lead times and finally better efficiency in plant**. Similarly Nair (2015) and Schoenherr & Speier-Pero (2015) suggest same advantages of analytics. Nevertheless, none of the articles go deeper into each of the benefits and how they can be achieved and what kind of data is necessary for them.

Procurement is in charge of upstream suppliers and global sourcing. The large scale of global sourcing results in massive number of transactions and it needs connection with internal finance for visibility, however a large proportion of company's cost is external expenditure. This expenditure is usually not consistently categorized and integrated with internal costs. One hurdle in integration of this data such as purchase orders (PO), shipping and invoices is their semi-structured nature. So data integration should not be limited to expenditure but it should cover the whole procurement (Rozados et al, 2014). Stakeholders need to know how their money is being spent. One use case of data analytics in procurement is prevention of fraud. Nevertheless, visibility gained through it gives better insights to stakeholders for developing sourcing strategy (Tan et al, 2015). Additionally "Analytics has, in the recent years, been used as a means to significantly reduce costs; by reducing the inventory turnover, transaction costs, cost due to wastage and also by reducing the time spent on negotiating for sourcing and supply purposes" (Karthik et al, 2015).

Chowdhary et al (2011) explains that firms might have visibility and transparency over procurement of core products but other product purchases are ad-hoc without monitoring. This is a source of spend leakage that goes unnoticed. Getting grasp of these unnoticed expenditure can help company's income in crisis time. They argue that current tools available for procurement are mostly reporting tools such as tools that process POs and filters them by different categories. However these tools cannot analyse data to find problematic areas.

Supplier management is an important aspect of procurement and identification of suppliers, monitoring their quality, performance and risks is an inseparable part of it. Furthermore, in strategic sourcing one company can opt to have multiple numbers of suppliers while another one tries to minimize number of suppliers. These decisions can be described by Kraljic's (1983) framework. Based on his framework complexity of supply base can make a supplier a strategic supplier that can have the advantage over the firm. On the other hand with lower complexity of supply base and how valuable a product is for firm, the firm can have leverage over the supplier with lower switching cost. If a firm seeks to change its position of power relative to a strategic supplier then it needs to find alternative suppliers to disrupt power balance and achieving better deals. Therefore, a firm's sourcing ability to finding alternative sources of supply becomes important not only in case of power position but also in risk management. Analytics aids in creating a benchmark of industry best practices to evaluate suppliers using multi criteria decision making techniques (Weng et al, 2016)

Mena et al (2014, p.94-95) discusses points to consider while adopting global sourcing. He argues that while companies can fully trust their immediate suppliers (first tier suppliers), they do not have transparency over their suppliers' suppliers. Consequently, he deems having a control tower to provide "bird-eye" monitoring necessary in controlling lead time and transparency of n-tier suppliers. Furthermore, he explains that in order to have a resilient supply chain that is guarded against supply disruption risks, supply chain should be mapped and critical paths of supply should be identified. Keith et al (2016, p.233-234) talks about gaining market intelligence by analysis of external market. It is argued that you need to understand what can cause disruptions in supply chain whether through something such as a marketing campaign or natural disaster. More importantly you have to be aware of flow of supply. Therefore, there should be awareness of location of suppliers and how they manage supply flow of their upstream suppliers. Even though having visibility of upstream supply can be valuable, they do not consider practicality and feasibility of such efforts.

2.5.1 Impact on procurement

Setting analytics to work for procurement requires thorough understanding of sourcing function. Traditionally sourcing was viewed as purchasing process. However, today we know that sourcing is a strategic function. Since substantial amount of a firm's revenue goes for sourcing, firms have realised it no longer can be a function operating in silo. Conversely it needs to be centralized and communicating with all other functions since its tasks are more than merely purchasing goods. Consequently, applying analytics to sourcing has become more complex. Sanders (2014, p.131-147) talks about impacts of Big Data on sourcing in following categories: order processing, standardization, visualization, more and faster queries, cost savings, predictive edge, enabling co-creation and risk management.

Perhaps the first thing that comes to mind when we are talking about Big Data and information technology (IT) is its application in **order processing** and automation. Speed of order processing affects customer satisfaction. Order processing is done through customer interface communicating with firm and Big Data improves its efficiency. Electronic data interchange (EDI) speeds up order placement and shipment and automation cuts excess costs of logistics, inventory and warehousing because it is monitored in real-time. Some companies like IBM have divided this cycle into two cycles, order to cash cycle and procure to pay (Sanders, 2014, p.131-147).

Having automation in place aids in **standardizing processes** and eases global sourcing. It transforms isolated parts of supply chain around the world into one integrated system that customer can contact any of the branches in any country and have its order fulfilled with same standards and procedures (Sanders, 2014, p.131-147).

Another use case for Big Data analytics is that it enables development of new applications and improving existing ones through its ability to **visualize** big databases and revealing relationships by using formats such as image, diagrams and animations. These aggregated or disaggregated data can further be combined with other data sources such as market information to yield more insights by what-if and scenario analysis (Sanders, 2014, p.131-147). Something that could be done manually and it was cumbersome since one could not collect all the data by his own and without visualization in an easy to understand format, it was difficult to make sense of data.

Big Data application enables the ability to run **fast queries** of different number of scenarios - Scenario analysis aids managers to form views of what might happen in future and what would be the optimal contingency plans to uncertainties of future events and markets (Fan et al, 2015) - additionally, it provides more dimension and filters to view data with lenses such as locations and categories. Furthermore, it facilitates having total cost of ownership (TCO) view on costs and procurement challenges and providing a competitive advantage by answering scenario questions (Sanders, 2014, p.131-147).

One of the reasons for having a centralized procurement and communication with other functions is collaboration to make decisions based on TCO. Procurement needs to be coordinated with designers and suppliers to **drive down and save costs**. Also collaborative planning, forecasting and replenishment (CPFR) is another mean for forecasting and coordination of different players. It is important to note that with Big Data application it is much easier to foster these collaborations (Sanders, 2014, p.131-147) and include more players and data sources.

Uncertainty is always concern of managers and they seek ways to overcome it. In terms of sourcing, uncertainty can stem from scarcity of supply, or change in performance and quality of top performing suppliers or change in commodities' price and currency rate fluctuations. Big Data tries to shed light on these issues by scenario planning and forecasting to provide a **predictive edge** to procurement (Sanders, 2014, p.131-147). Teece et al (1997) defines dynamic capabilities as ability to reconfigure resources based on changing environment. Building on this definition, Pavlou and El Sawy (2011) propose four dynamic capabilities as tools for resource reconfiguration. These capabilities are sensing, learning, integration and coordination. Improvement in forecasting can act as the sensing capability towards external world. This capability is also beneficial in risk management. Consequently, Big Data analytics can help in building dynamic capabilities. Chen et al (2015) mentions that information processing is a dynamic capability that provides temporary competitive advantage, however Big Data analytics is more than simply information processing.

Nowadays one of the hot topics in business models is a movement to service dominant logic and creating superior value for customer. Proponents of this movement believe that superior value is achieved through **co-creation**. As an example for manufacturing huge products that require great number of suppliers, Big Data provides the ability of sharing information with the whole chain so that various companies design in accordance with each other. Consequently,

with co-creation not only end product is of higher quality and lower defects, it saves time and costs (Sanders, 2014, p.131-147).

Sanders (2014, p.131-147) explains that analytics for **risk management** are underdeveloped even though they have great importance. Instead of simple metrics he suggests supplier resiliency score. This score considers some numbers risk factors for suppliers and then shows if a specific supplier needs to be monitored for risk and contingency plans. As an example of risks associated with this number, facility location and availability of alternative sites of production is taken into account. Wang et al (2016) suggests that publicly available news and social media channels related to suppliers and markets are sources for risk identification. Another area for risk management is assessing supplier's performance in response to the risks they have encountered in case of economic crisis or scarcity of suppliers and other relevant events. Weng et al (2016) believes that analytics can help in monitoring performance of suppliers by comprehensive supplier's data collection and quick analysis.

Another area to consider in Sanders (2014, p.131-147), is child **labour and human trafficking**. There are strict legislation in some countries regarding this issue and if a supply chain is operating globally then it becomes an important issue. Legislation and laws not only can disrupt the supply chain but also human trafficking damages company's image and reputation. Big Data analytics helps in identifying high risk areas and detection of this issue. Big Data can utilize discussions on social media to give warnings about potential risk related to forced labour and then firm can investigate issue through more reliable sources. It is important to note that with increasing trend of outsourcing and relying on suppliers, risks associated with **supply disruptions** surge. Furthermore, by adopting lean production and having minimum inventory, a shortage can disrupt a supply chain. To have a shield against these uncertain events, predictive analytics and scenario planning can become handy. It is important to identify critical paths of supply chain and predict what can go wrong and having a backup solution for that.

2.5.2 Potential sources of data

Fan et al (2015) suggests a framework for managing risks that data should be considered both internally and externally. Internal data are purchasing records for checking quality, bill of material (BOM), delivery records, financial records, crucial facility records and human resources (HR) records. For external resources following sources are suggested: public news, politics, weather records, natural disaster records and social media.

The type of social media analysis that is influential for business is social influence analysis. In this type of analysis, influence and their connection in social network are modelled and evaluated. Since behaviour of actors is affected by other actors, it is valuable to understand patterns and strength of influence (Gandomi et al, 2015).

Furthermore, TATA consultancy's report by Chithur (2014) suggests **experience of category managers, consultants, industry associations, trade journals, company annual report and online literature as useful sources of data**.

2.5.3 Type of analytics needed

The type of chosen analytics depends on how complex a supply chain is. Supply chains can either be stable, established with low uncertainty or can be evolving with ever changing technology and supply and demand fluctuations. Sanders (2014, p.131-147) determines supply chain complexity by comparing two aspects, demand uncertainty and supply uncertainty and defines four types of supply chain:

- Efficiency focused SC (low supply uncertainty, low demand uncertainty)
- Responsive SC (low supply uncertainty, high demand uncertainty)
- Risk-Hedging SC (high supply uncertainty, low demand uncertainty)
- Agile SC (high supply uncertainty, high demand uncertainty)

Efficiency focused supply chains are easiest to manage and predict and the eliminating waste improves the benefits. Therefore, for this supply chain automation of order processing and a simple visualizing dashboard suffices.

Responsive supply chains are used for innovative products. Because of low uncertainty of supply we are mostly concerned with demand uncertainty. Consequently postponement principle is adopted for this SC. Examples are Zara and Hewlett-Packard (HP) who postpone final operations until demand becomes clear. For this SC analytics should be used for implementing postponement.

Risk-Hedging supply chains are prone to supply disruptions and inventory shortage and they keep high inventory stock and share their resources between different companies. For such situation scenario planning comes handy.

Agile supply chains are hardest to manage as they should cope with high customization on demand side while preventing supply disruptions on demand side. Again similar to risk-

hedging supply chains they need higher safety stock and they pool and share resources. For this situation, predictive analytics and scenario planning are necessary to avoid risks.

Karthik et al (2015) conducted a study to understand relationship between level of analytics adoption and its perceived benefits in procurement. By level of adoption they mean the degree that analytics cover functions and their perceived benefits are financial benefits, inventory benefits and supplier intelligence. Financial benefits are related to procurement costs, price reduction and customer satisfaction. They found out that level of adoption has direct impact on financial benefits. Maximum benefits is gained when all the departments have put analytics in use. For inventory benefits, the more level of adoption the more benefits gained. This is due to quick decision making for supplier selection and inventory optimization. Despite these benefits, the interesting finding was the relationship between supplier intelligence and perceived benefits. With increase in degree of adoption, supplier intelligence tends to decrease. They argue that this is due to transparency becoming a routine and relationship with supplier becomes the hindrance. When all suppliers are participating, price competition becomes a problem to perceived benefits. In addition to that, relationships and supplier switching costs are problematic. However, overall benefits gained is increased by adoption of analytics.

Another factor that should be taken into account for implementing analytics is change management. As mentioned earlier in framework of business analytics, there is a need for corporate wide movement and cultural change to buy-in analytics and trust data. Sourcing is no different in that sense. Richey et al (2016) mentions answer of one supply chain professional regarding this issue that Koreans are not familiar with decision making based on quantitative. Their past experience and opinion of people whom they trust is more important when it comes to decision making. The obstacle is **emotional attribute**.

2.5.4 Current state of procurement and Big Data analytics

One place to start observing the situation is spend analysis. Correct spend analysis can increase visibility and identify cost saving opportunities. Nevertheless, success of it is largely reliant on the ability to appropriately access, organize, and analyze data as mentioned by Limberakis (2012). Spend analysis is a top priority operation for procurement professionals that requires integration of multiple data sources and clearly defined data processes. Also procurement needs coordination of other functions to achieve success. In research done by Limberakis (2012) the core issue of spend analysis was **quality of internal data** from ERP and procurement systems

that hindered correct forecasting and identification of saving opportunities. Data quality problem comes from multiple data sources that are not integrated and from usage of manual spreadsheets. 32% of respondents of that research were using spreadsheets as their primary tool. A possible solution to this situation is **automation** from extraction of data to its enrichment. Nonetheless even in firms who are more advanced in analytics, data extraction has the highest level of automation whereas cleansing, classification, and enrichment are mostly done manually.

Limberakis (2012) suggest that the first step in creating more effective spend analysis tool starts in data extraction effort. The issue of data extraction from multiple sources and cleansing data should be solved and become automated. Then company-wide standards and formats should be adopted so make common understanding across company. In the transformation process suggested by Limberakis (2012), issues need to be solved internally at first and in the last stage company can make use of external data such as market indices, commodity prices and external events impact to be able to make use of predictive analytics and manage risk better. But spend analysis is only one part of procurement.

Rafati and Poels (2015) argue that for fact based decision making, a firm needs both data management and data analytics competency. Data management is more concerned with technical side of issue such as transformation, storage, governance and etc. Data analytics is the part that can derive value from data which was discussed earlier in this review. The situation with procurement analytics is that they use analytical techniques to aid decision making, however the burden is data management capability. The challenges that firms face in managing their data are **inaccurate and not timely information of spend patterns and suppliers**. In addition to that there are vast volume of data in **dispersed systems** that makes its management even harder.

To understand procurement analytical layers and their challenges Rafati and Poels (2015) made the taxonomy in Figure 4. The first layer is application layer. Here lies various systems such as ERP, e-procurement, request for (RFx). The challenge with this layer is that procurement data is scattered across different systems that are not integrated. Next layer is process layer that has key sourcing activities such as spend management and supplier selection. The problem with this layer is that not all of these processes are application based and if there is data for analysis, it is not in digital form. The third layer and the core layer is data layer which contains data on

contracts, suppliers and spend. There is lack of platform to consolidate all data sources from pervious layer and a shared storage for analysis.

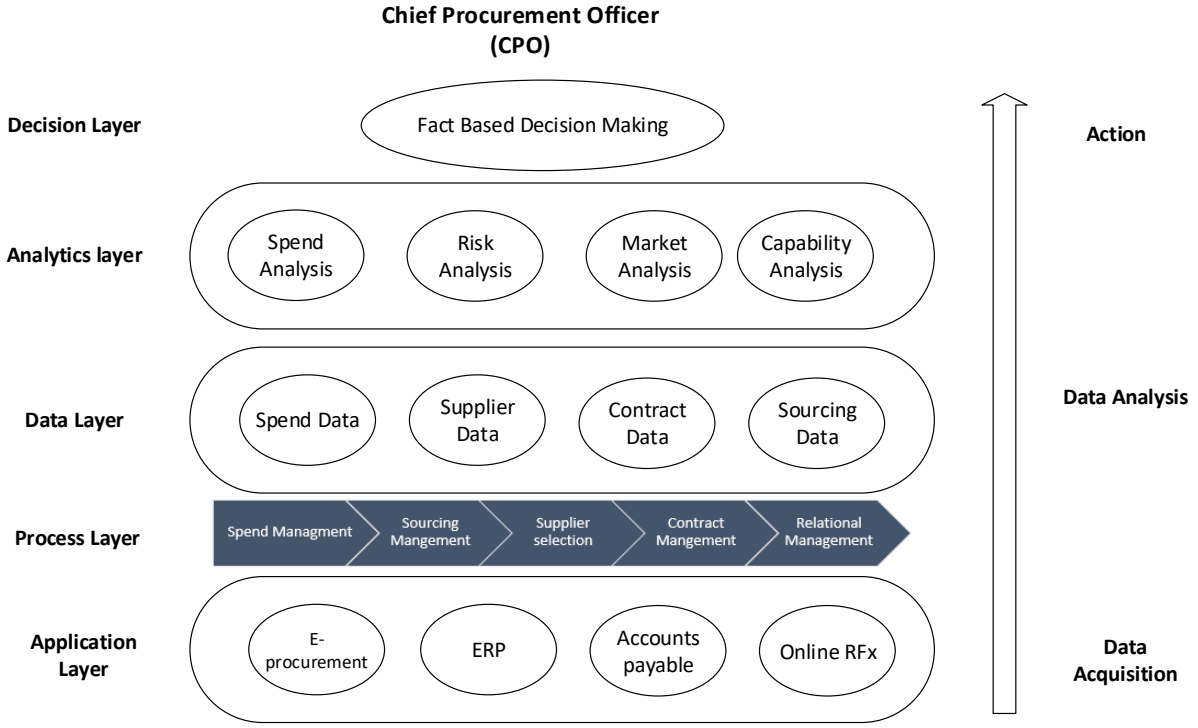

Figure 4 Layers of procurement (modified from Rafati & Poels, 2015)

The fourth layer is analytical layer that consists of tools such as market analysis and cost – benefit analysis. This is the layer that derives insight from gathered data to add value to business. Depending on maturity of organization and type of analytics in use, they might lack tools and skills to make use of data. Final layer is decision layer. In this layer there is a need for visualization techniques to transform insights into understandable format for ease of decision making.

2.6 Literature synthesis

In this section a picture of literature review is depicted to have a comprehensive view of literature. Table 1 shows Big Data characteristics and analytics, Table 2 shows role of Big Data in supply chain and Table 3 shows Big Data in procurement.

Table 1 Synthesis of Big Data and analytics

Big Data Attributes	Type of data	Type of data analytics
• Volume • Velocity • Variety • Veracity • Value	• Structured (least volume) • Semi-structured • Unstructured (highest volume)	• Descriptive • Predictive • Prescriptive

Table 2 Potential of Big Data in supply chain

Big Data in supply chain		
Key success factors	**Benefits**	**Challenges**
• SC system integration • Improved forecasting/decision making • Human capital • Security and governance • Storage • Operational efficiency • One version of truth	• Increased visibility • Enhanced bargaining position • Ability to respond faster to change • Greater power in relationship with suppliers and customers • Operational efficiency • Accuracy of information • Better risk management • Informed decision making	• Needle in haystack (where to start?) • Islands of excellence (uncoordinated functions) • Measurement minutiae (too many metrics) • Analysis paralysis • Data quality • Dispersed systems • Lack of standard data format

The success factors, benefits and challenges of Big Data in supply chain is not completely different from Big Data in procurement. In fact supply chain analytics is a bigger version of analytics. For example, SC system integration, one version of truth, problem with data quality, dispersed systems and some other factors are also true in case of Big Data in procurement. The Big Data in procurement has a more detailed view of analytics used in procurement but has similarities to supply chain analytics. Therefore, the efforts needed to succeed are similar to each other. This view of literature exhibits that literature explains general issues related to Big Data analytic and do not deep further into them to understand how they can be implemented and what their hurdles are, especially in procurement.

Table 3 illustrates the impacts that Big Data can bring to procurement such as standardization of processes, driving down costs and providing predictive edge. Internal and external sources of data are also compiled. In the two last column of table procurement layers alongside their challenges are depicted.

Table 3 Big Data in procurement

<table>
<tr><th colspan="5">Big Data in Procurement</th></tr>
<tr><th>Impacts</th><th colspan="2">Potential sources of data</th><th colspan="2">Procurement layers and challenges</th></tr>
<tr><td rowspan="5">Order processing
Standardizing processes
Visualization
Fast queries
Drive down and save costs
Predictive edge
Co-creation
Risk management
Supplier evaluation</td><td>Internal</td><td>External</td><td>Decision (highest layer)</td><td>Need for visualization techniques</td></tr>
<tr><td rowspan="4">Purchasing records
BOM
Delivery record
Financial Record
Crucial facility record
HR records
Experience of category managers</td><td rowspan="4">Public news
Politics
Weather records
Natural disaster
Social media
Consultants
Industry association
Trade journals
Online literature</td><td>Analytics layer</td><td>Lacking tools and skill to make use of data</td></tr>
<tr><td>Data layer</td><td>Lack of platform for consolidation of data from previous layers</td></tr>
<tr><td>Process Layer</td><td>Not fully supported by applications
Data not available in digital format</td></tr>
<tr><td>Application layer (lowest layer)</td><td>Scattered data</td></tr>
</table>

3 Research methodology

In this chapter, qualitative research method in form of case study and the reason for its selection is discussed. The phenomenon we are addressing in this master's thesis has not been researched extensively in literature. Therefore, the method that helps to explore new aspects of this phenomenon is qualitative method. The advanatage of qualitative research is that it provides possibility of in-depth study of various topics (Yin 2015, p.6). Qualitative research gives us the ability to study people without constraints of other research methods such as questionnaire and it enables discussion between interviewer and interviewee in form of semi-structured interviews. This purpose of this study is to analyse use cases of exploiting Big Data for purchasing and supply purposes, in order to help companies in having more visibility over the supply market. Consequently, with qualitative approach we are able to capture their perspective on the issue and since it is an explorative study, discussion with interviewees were beneficial in enriching mind-set of researcher while opening further avenues for question.

Qualitative study can be conducted in various forms such as grounded theory, ethnography, critical theory phenomenology and etc. However, case study was chosen to address the research gap around this topic. Case study allows us to have more than one source of data. Baxter & Jack (2008) believe that case study explores research topics with various lenses and it can capture and reveal more aspects of the phenomenon under study.

3.1 Case selection and data collection

For case company selection we were searching for companies that are at least advanced in their purchasing and supply management and have visibility over their expenditure. The reason behind that to target companies that have already realized importance of sourcing function and were willing to find new tools for its improvement. Additionally, we were considering companies that had adopted category management principles. We assumed that each category has its own unique challenges and we can capture various insights with different categories. Nevertheless, we prioritized strategic categories necessary for business regardless of being indirect or direct categories. There was no preference for any specific industry, service providers or manufacturing companies. In our search for case companies, we used assistance of one procurement software provider to identify their customers that matched our criteria. After examining companies, six case companies were selected. These companies represented

two engineering companies, one from aviation industry, one from telecommunication industry, one from food industry and one from industrial machinery. Upon request of case companies their names have been anonymized.

Data collection was done in form of semi-structured interviews. There was an iterative approach to interviews as interviews were conducted, new issues were raised that had to be addressed in interview questions. Overall, nine interviews were conducted. The main respondents to interviews were chief procurement officers (CPO) but category managers were interviewed where it was possible which added more value. CPOs holistic view to the topic whereas category managers saw topic in greater detail since it was their main responsibility.

Interviews consisted of four main parts. In the first part, the interviewee talked about his/her background and his/tenure in company. In the second part, managing principles of the chosen strategic category were discussed, what makes the category important and what challenges are in managing the category. This was necessary to have a clear picture of how the procurement is acting in the firm. In the third part, advanced analytics situation of the company was addressed. Finally in the fourth part benefits and risks associated with Big Data analytics were discussed.

Table 4 Interviews

Case number	Case company	Title	Duration
1	Engineering 1	Head of indirect sourcing	51:34
2	Engineering 2	CPO	55:00
		Global cluster manager	29:15
3	Aviation	CPO	60:00
		Global head of indirect sourcing	38:32
		Category Specialist	01:03:01
4	Telecommunication	Head of innovation collaboration ecosystem	55:14
5	Food	Senior procurement manager	52:57
6	Industrial Machinery	Senior vice president (SVP) of global sourcing	30:48

Table 4 summarizes interviews. Four out of nine interviews were conducted face to face and the rest done through Skype or phone call. All the interviews were held in English. With

interviewees' consent, interviews were tape recorded and transcribed solely for research purposes except for one interview that its data was lost.

3.2 Data analysis

Analysis of empirical data was done through coding transcribed interviews with Excel. Some of the codes were predefined, they were taken from literature review and some of them emerged in coding process itself. In order to facilitate representing results eight main codes were used and some of them had their own sub codes.

The first code was ***category management*** that collected background of procurement function and its challenges. This directed interviews towards the challenges of that specific category. The second code was ***supply base***. This code compiled data on ways of identification of new suppliers, monitoring them and importance having visibility over n-tier suppliers. The third code was ***data analytics current situation*** that addressed type of analytics in use, tools, sources of data and current situation with its shortcomings.

The fourth and fifth code were ***desired state*** and ***purpose of use of Big Data analytics***. Due to overlapping data between these codes they were considered under one code, ***desired state of Big Data analytics.*** This code addressed expectations of interviewees from Big Data analytics and how it can help in supply risk management. The sixth code was ***challenges and risks of data analytics***. The seventh code was ***representation of data*** and the final code was ***others*** that was about role of social media and trusting data analytics.

Interviews were analysed one by one and coded simultaneously. After coding data pertaining to a specific code and its relevant quotation were grouped with each other in an Excel file that complied all of the codes and their data. This helps in comparison for finding similarities and differences between case companies and to understand how each category can make differences. The results were obtained by compiling similar data with similar codes and grouping them. The codes were compared across cases so that similar patterns could be found.

4 Analysis and results

In this chapter the findings of the study are presented. First, supply base of companies are explained with its special attributes to see if there is room for improvement through Big Data. Second, current state of data analytics in companies are presented, what they are using, what is working and what is challenging. Next, expectations and optimal state of data analytics are studied and finally importance of social media, trusting data analytics and risks of Big Data are discussed.

Before going further into results, some information about case companies are presented in Table 5.

Table 5 Description of case companies

Case company	Category	Description
Engineering 1	IT	• Management of all software and hardware • Time allocation for supplier management is challenging • Unique cost challenge • Digital transformation requires investment • Keeping the budget is challenging
Engineering 2	Direct sourcing	• Project business and small scale car part manufacturing • Challenge of information transparency across all locations and across all parallel projects • Challenge of supplier related data availability, supplier basic data and operational data • Challenges in on time delivery, KPI information across different ERP systems

	Electrical parts	• A functional team with different business representative for identifying strategy and supplier selection • Heterogeneous category
Aviation	Travel	• All displacements of staff for business purposes whether through air, rail or vehicle transportation along with accommodation • Operation in 46 countries with 46 different ways is challenging • A lot of detective work
	Equipment	• Sourcing equipment around globe • Tendering annual volume • Obtaining quality forecast of next year equipment is challenging • No good quality data on needs of different stations • Emotional matter in switching suppliers
Telecommunication	Software	• Movement from hardware to software and virtualized solutions with different pricing model • More about relationship building • Need of mature solution for long term use, at the same time, you have to be agile to find the best solution
Food	Raw material	• Bakery and confectionary in different countries. • Keeping right amount of stock is challenging • Price is a really big issue and there are some material which are heavily

		influenced by market situations, market prices. • Big raw materials like sugar, coco and flowers can be managed better than long chain of small materials. • Problem with availability of tailor made raw material with high prices that require accurate prognosis • Food fraud is a big issue in the future • Need to find a way to work with suppliers to get same figures
Industrial Machinery	Casting and Forging	• Critical parts of all machinery that are manufactured • Longest lead time, highest share of product cost • Long time to develop suppliers • Significant quality risks

4.1 Supply base

In this section supplier identification methods and necessity of having visibility on n-tier suppliers are discussed.

4.1.1 Supplier identification

In category section of interview where the purpose was gaining background of current sourcing practices, supplier identification issue was raised. It is important to note that identification methods varied between firms due to unique characteristics of each category. Additionally, it proved that basing study in category management yielded in broader views. Methods of identification were as follows:

Case 1 engineering 1, IT category:

This firm relied on a consultancy to get list of suppliers and colleagues in other companies (**Reference**)

Case 2 engineering 2, direct sourcing and electrical parts category

This firm have project business and small scale manufacturing. However, its supply base comprises of well-established corporations, medium and small size companies. The nature of project business affects supplier selection as described in interview CPO of case 2 engineering 2:

"*As we are delivering production equipment there is always certain customer specific standard and customer specific approved vendor list that we need to apply, which means that in some projects, some of the suppliers are* ***pre-defined by our customer****. "we do normal sourcing work, which means that we use different sources like* ***industry fairs, internet, industry association*** *and so on, to build supplier long list. Industry segment is not very wide so there is good knowledge existing" (CPO case 2 engineering 2)*

Case 3 aviation, equipment and travel categories

Most important way of identification in equipment category is **conferences and industry fairs.** This method of identification is dependent on category that was discussed. It was used in aviation industry, one engineering company and food industry.

In aviation industry, equipment category suppliers were identified in annual conferences held in Las Vegas and Munich. In these industry fairs suppliers gather to show their innovations to their prospect customers but the reason that this category is relying on fairs is that aviation industry is regulated for safety and for environmental impacts. Identification of suppliers through other sources such as public data cannot guarantee that a suppliers passes regulation. Category specialist of case 3 aviation describes that:

"In these conferences ***supplier has gone through due diligence*** *and has checked all the requirements and fulfils these requirements in order to be a worthy supplier. We believe that the right place where all such recognized and conforming suppliers are uniting is during these conferences. The aviation industry due to its nature is not that large, people tend to know each other in between the competitors, in between manufactures" (category specialist case 3 aviation)*

In travel category, the target is mostly airlines and their identification is not challenging.

Case 4 telecommunication, software category

The category discussed in this interview was software category. It was quite distinct from other categories which were more or less manufacturing related. Looking specifically at software, supply base can consist of small start-up teams and managing them is mostly important from relationship management perspective. To identify new suppliers they rely on **scouting** done by engineers as described by supplier relationship manager of case 4 telecommunication:

"So that we can build shared knowledge cross business groups and business lines so that they would be combined knowledge built of market situation and especially category strategy work. This combined knowledge is then combined with market studies" (Head of innovation collaboration ecosystem case 4 telecommunication)

Case 5 food, raw material category

In food industry there is also a strict supplier approval process that suppliers need to be tested in a lengthy process until they are accepted for business. Identification of suppliers is done by sourcing managers and product development teams by **asking suppliers or meeting them in exhibitions**.

Case 6 industrial machinery, casting and forging category

Identification of suppliers is kick started by a sourcing project. Suppliers are identified through **web** and **peers**. However, it was mentioned that they are large number of suppliers in world that they are not sure whether they are using best ones or missing out on someone.

Out of methods of identification, the most common used method was relying on references or industry fairs depending on needs of that specific category. However, none of the companies were using a public source of data to get access to more potential suppliers, even though they were using internet to find more options, internet is not a trustworthy source for all companies, as category specialist from case 3 aviation mentions :

"For big data, there is a point where any given person is a student whether it is for bachelor's degree or master's degree where you believe that internet is helpful in terms of knowledge. Wherever you start working and that's important this knowledge becomes of insufficient quality in order to make real decisions in order to motivate stakeholders to change their strategy, to change their tactics, to redirect budgets" (category specialist case 3 aviation)

4.1.2 Transparency over n-tier suppliers

Considering advances in technology and emergence of IoT, tracking a supply chain has become easier. Therefore, the issue of monitoring n-tier suppliers and its importance was addressed during interviews. The goal was to understand whether it is important to have transparency to more than first tier suppliers or not.

Not all companies saw the issue similarly. Again the category discussed was influencing their insights. For travel category manager, n-tier suppliers were not a concern, he was more looking into managing internal users of category. Additionally, in case of software category there were not much of suppliers of suppliers to consider, however with cloud and SaaS, there are stacks underneath that software is built on which is worth monitoring. In case of casting and forging category, mapping supply base was easy because there is not a long supply chain after casting supplier.

The rest of categories saw importance in visibility over n-tier suppliers to some extent. One concern was understanding routes of supply and also risk management, senior procurement manager of case 5 food puts it this way:

"Nowadays if you think backwards to suppliers and then there is a big really many that kind of steps backward so that there is a long supply chain. Then of course you need to trace your supply chain to that direction and also manage the risk, what can happen. I think ***the risk management is a key issue*** *that we need to know how to follow the risk in those long supply chain and where to get the information" (senior procurement manager case 5 food).*

Another important factor that was mentioned in interviews was **ethical and corporate responsibility**. In case of food industry, customers were concerned about where the raw material was coming from. Not only there are ethical issues, in some countries there are **legal requirements** concerning them. Legal labour is another aspect that one company was concerned regarding importing labour from china.

Electrics and electronics cluster manager was very eager to have such transparency. In electrical components there is a lot of subcontracting that subcontractor is purchaser of materials and there is no visibility over that. There are also benefits in negotiation with direct

suppliers if there is awareness of what is happening behind the scene to gain **better position in negotiations**.

Regarding the practicality of transparency over n-tier suppliers, some suppliers are comfortable with working through emails as the common practice but implementing something advanced on top of it makes them concerned. Change management and willingness of suppliers play important role here as data is preferred to be provided by suppliers themselves. It was mentioned in interviewees that there is a need of having direct relationship with 2nd or 3rd tier supplier since public data is not reliable. Besides that having a system to system connection requires heavy investments and capabilities of customer and supplier for interconnection are not always equal. To summarize, supply chain tracking in terms of **supply and legal risk and gaining leverage in price negotiations** are the most promising use cases of transparency over n-tier suppliers.

4.2 Data analytics current state

Out of the 3 types of analytics (descriptive, predictive and prescriptive), all of the firms had basic descriptive or reactive analytics however only two of them had predication and none of them used prescriptive analytics. Some use predictions in their discussions but not with a systematic and with technological tool. Tools used by the firms were spend management, contract management, artificial intelligence (AI) and supplier data base tools. Nevertheless, firms were not quite satisfied with current state of their systems. In some firms such analytics are only spend visibility tool that looks at averages, minimum, maximum, spend per region and fraction of spend.

Hurdles that firms are facing can stem from state of internal data. For better classification of problems in this area, they are presented in 2 main problem area: quality issues and manual work.

When we talk about quality of data, we are talking about quality of decision making as well. Almost all of the firms were worried about “garbage in, garbage out”. Inaccurate data leads to wrong decisions. It also affects trusting data analytics. While algorithms might function flawlessly, with wrong data there would be no useful outcome. There are some reasons that quality of data suffers. Companies usually have standalone reporting systems with different data that are conflicting. One firm reported that their prognosis, their estimates and their budget figures are deviating from each other and it makes it difficult to decide which number should

be linked to volumes. In some firms there is no robust data collection practices since there was no expectation of current capabilities that data analysis can provide to an organization. Also building a system with considerations regarding data analytics is much easier than rectifying current legacy systems.

The travel category manager case 3 aviation pictures situation of standalone systems this way:

"We are in 46 countries and we have 46 different ways of doing things. There is no standard general ledger, we report things like I said in 46 different ways" (global head of indirect sourcing case 3 aviation)

Some of the systems that are implemented, are not implemented in all parts of globe and not all functions are making use of them. So in these areas information need to be pulled from account payable which there is problem of misinterpretation and misclassification. Furthermore, accessibility to supplier base is not equal around globe. So a supplier might be known to one local procurement while in another area of world, procurement do not know this supplier and can't have good positioning in negotiations. The interesting point is that even when procurement is centrally managed, there is still this incoordination among procurement functions. Even if data in procurement function is harmonized, procurement is no longer operating in silo and it needs data from other functions, however data taken from other functions might be in an undesirable format.

In companies that recently have understood importance of procurement and are moving towards centralized procurement, there is no prior data, still there are many local and ad-hoc expenditure that makes it hard to derive insights from data. Another issue is different practices of data storage and sharing both in functions other than procurement and functions that operate in different location. They may devise different categories or use different abbreviation and conventions.

"I can lose up to 50% of my data just through ***misclassification and miscategorization*** *and currently I don't have the staff I don't have the analyst in place" (global head of indirect sourcing case 3 aviation)*

*"****Quality is probably the biggest hurdle*** *and there is for example* ***no common coding for components, all the entities code their components themselves****." (global cluster manager case 2 engineering 2)*

"There are a lot of acronyms here, I think that is one of the things that has been bothering mostly. Every department uses this abbreviations two to three letters and you have no clue what they mean and you spend huge amount of time just trying to get what on earth this might mean" (Head of innovation collaboration ecosystem case 4 telecommunication)

Another problem with dispersed data is gaining insights from them. One of the companies mentioned that it is easy to identify main suppliers and purchases but then there are purchases that are made in other locations that we do not know why they made that purchase and what kind of supplier we are dealing with, whether it is a long term supplier or just one off need supplier. There is good visibility over invoices and POs but no visibility over who and why. This hinders cost optimization and saving opportunities.

"I mean right now I can have spent in billion dollars but if it is ***scattered*** *and I don't understand what is in there, that is a* ***loss of money and loss of opportunity****" (global head of indirect sourcing case 3 aviation)*

The second problematic area is manual work. Not all the data is system supported and those that are not supported are manually collected through Excel. Excel was mentioned many times during interviews and seems as a common tool among companies. Some companies have manual reporting system which is usually done through Excel. The data of past can be presented and plans can be reviewed with head of different departments, however these are done manually. One firm believed that Excel is their most important tool even though it is manual and was mostly concerned about getting things done. But this was not the case for another firm which was concerned about resource intensity of manual work:

"we have a lot of manual work in that and internally we lose a lot of resources to analyse in different kind of data and huge Excels so that is the problem in internal data, we can get and we can discuss, we can develop that but it is huge amount of work" (senior procurement manager case 5 food)

Updating supplier related information is done manually. One firm explained that there might be procurement plans that are shared among different projects. These information might not be fully supported by system so there is no online transparency over them and to access information, individual people who have access to project specific folder are asked to provide data. In some cases there are manual calculations in purchasing and controlling. In addition to manual reporting or manual calculations, compiling data from different sources is manual and demanding. For instance one firm was using score cards to rate its supplier, however they had

to get information from two different systems and connecting them together manually. Even in mapping supply base the challenge is not always getting information but the manual work needed for it. Additionally, systems that are not fully implemented around the globe need manual effort for coordination.

*"There are multiple sources and currently **very challenging to combine sources for more sophisticated cross analysis** except if it is manually done which takes effort and workforce. Data amounts are so massive that you always have suppliers that are outside the system and you still need manual effort to ensure **data accuracy**" (CPO case 2 engineering 2)*

Table 6 Internal data challenges

Quality related issues	**Manual work issues**
Garbage in, garbage out Misinterpretation, classification and coding problem Different format and naming conventions Deviating internal data	Manually collected in Excel sheets Challenge to compile multiple sources of data for analysis Manual effort for ensuring accuracy Resource intensive

Table 6 summarizes challenges of internal data. Standardization can be a solution to quality issues and facilitate automation of processes. Nevertheless, standardization needs huge investment, justifiable business case and return on investment consideration. One interviewee believed that most of companies do not consider indirect workload as part of business so it has less priority than solving business related or ad hoc matters.

""It is almost impossible to explain the benefits of completing this task (standardizing data) to top management as to justify existence in company" (category specialist case 3 aviation)

Additionally, it was believed that standardization is a large task that a company cannot do it on its own and usually consulting firms undergo such practices. In case of outsourcing standardization, requirements and format must be set to avoid quality issues. Furthermore, it is important to have advanced expertise in domain knowledge to make sense of data and grouping it.

4.3 Desired state

In this part, expectations of supply professionals regarding Big Data are explained. The desired state of data analytics is seen from seven perspective, transparency, automation, prediction, data sources, new KPIs, representation of data and other relevant issue. Finally risk management is explained as a distinct use case of Big Data analytics.

Regarding transparency one firm was concerned to have a system in place that is able to building a big picture while filtering human's opinion. Another expectation was presence of a central system that contains **supplier related information, procurement related information and supplier performance information**. Additionally, this system can **contain market related information in terms of macroeconomic, industry segment and who are the main players**. It is beneficial if this system has **transparency on other industry or technology segments** and can find global products and suppliers. This makes sourcing efficient by not relying on catalogues. Going deeper into what information of suppliers is necessary to have one interviewee believed:

"There needs to be transparency to all the qualifications, supplier audit, financial check, risk assessment, action plans, relationship plan, whatever comes with running the relationship and the business together" (CPO case 2 engineering 2)

Moreover that system should keep record of analysis already done on suppliers, opinions, contracts and basic supplier information that should be kept up-to-date by supplier. Spend information should have transparency in a way that is **detailed and granular enough** to provide possibility for sophisticated analysis. Finally it is useful to have reliable data that is easy to search, quick to read and real time database facilitates searching and looking from different perspectives.

Next expected issue is automation. One company expected to have a way to easily identify more potential suppliers in accurate way. Case company 1 was concerned about identifying top of iceberg but not seeing beneath it and they could not identify who their long term critical suppliers are and who are non-critical ones. Such assessment was done manually that they expected to reach this understanding of supply base more efficiently. Another point mentioned by them was ability to gain information needed by one click, however it was pointed out that with Google you can get millions of pages of information but **relevancy is important and system should be able to filter irrelevant data**.

Usually data compilation was done manually especially if it was done among different systems, so availability of a tool for data collection speeds up the process. One firm was moving towards a software integration within the network that could utilize data from various departments to optimize cost savings. Nevertheless, it is also important to have detailed view of compiled data as expected by one interviewee:

*"I think the ability to **distil** big data down into **specific areas** to pull it out either with a tool like crystal reports or some other reporting tool allows us to **specifically target areas that we want to either negotiate or improve on our current position** " (global head of indirect sourcing case 3 aviation)*

One interesting expectation was predication ability. This issue was discussed in interviewees based on their own perspective and its benefits for their categories. For example, travel category manager expected to be able to identify holiday sessions and predict travel pattern if there is going to be an event in future and act accordingly. He expected to target law cost targets based on predication which could make 50% difference in travel expenditure. Another firm saw huge strategic opportunity in strategic sourcing gained by optimizing storage levels, forecasts and drawing conclusions based on them. To explain further this issue, it was explained that there is a need for using regressions and statistical distributions to drive decision making not only for spend but with TCO perspective.

*"I would like to have more forward looking, everything is looking to history. From **delivery capacity perspective or cost development perspective**. I think those two are main things. Knowing what the performance was then what the performance is going to be next month and so on" (SVP of global sourcing case 6 industrial machinery)*

A good question is what are potential sources of data and what can help supply professionals in decision making. As mentioned earlier one of interviewees believed that internet is not a good source of information. Instead he was in favour of academic literature and data available in peer reviewed articles. There are people who have already worked on the issue, so he considered **academic literature** as a starting point. He also suggested **international organization's literature** provided that you know what you are searching for, otherwise there are lots of literature. Interacting with not only suppliers but competitors in industry fairs is another potential source of data in his views. Other sources that are beneficial for decision making are supplier market share, supplier businesses and their pricing. Price benchmarks is

useful especially in direct purchasing where information is typically not revealed. Transparent pricing, data on quality and of course competition situation is of value.

"If you can capture from public databases information about who has been connected to whom and if you learn to use it kind of efficiently then you might see that for example competitor are dealing with those and those companies, maybe it's better to have a look at" (Head of innovation collaboration ecosystem case 4 software)

*"**Of course we maybe need some benchmarking, what the other ones are doing**. I think that main thing is to get benchmark information and early signs from that information, how the markets are changing and what is happening in supplier world that some suppliers are increasing and some decreasing**, also category based what are the benchmark prices**" (senior procurement manager case 5 food)*

An interesting insight that appeared in interview with the case company 4was that recordings of meeting was one potential source of data which can be exploited by speech analysis capabilities.

One company believed there is a need for other performance metrics to be easily usable not only on time delivery because there are different problems with suppliers that requires new **performance KPI**s.

"Currently many companies can only measure on time delivery and quality related information and then do for example annually some manual evaluation to asking people how they feel about the technology or technology position or service level and use this information with supplier's feedback"(CPO case 2 engineering 2)

Additionally, another company was looking for information to measure **supplier's flexibility**. It was argued that they need to measure how flexible is a supplier towards its customer's changes like production plan changes and how they react to changes. So it would be beneficial to have a new KPI on measuring flexibility and rating suppliers based on that. This can also support idea of another manager that Big Data can help in supplier quality management as well by real time assessment of performance and helping them in improving quality, punctuality and productivity.

One issue that was raised during interviews was that what representation of data they like to interact with. In case of present standalone systems, it is ideal to integrate these platforms and presentations. Additionally, there should be one set of numbers, **one set of reporting and one**

set of future understanding. One respondent claimed that are there a lot of excels and graphics but something more technologically advanced is needed. Obviously, all the formats should be available with a drilldown to show sources of data. Perhaps the best description was made by case company 4 that had already seen such idea in a start-up:

*"It was a **visualizing system** that was the cool thing. It's **quick to understand because you don't need to read, you just look at pictures and you can move and look at from different angles**, I think this is really cool" (Head of innovation collaboration ecosystem case 4 software)*

There are other points worth mentioning in desired state. Accessibility is important especially if system is implemented in different countries and it should be timely, so correct data at the right time to the right person. Big Data analytics should not be limited to spend but helps in decision making by supporting components of TCO. Big Data is needed to accumulate what the costs are per supplier, equipment and category on grand global scale. One manager mentioned that in optimal case there would be an **extranet with suppliers** that they can see what is shared with them and sort of one place to check information. Finally, transparency provided by Big Data analytics can bring stakeholders on board. Table 7 Expectations from Big Data AnalyticsTable 7 and Table 8 summarizes expectations of Big Data analytics.

Risk management

Risk management seems as a promising field that Big Data can contribute to it. One important aspect of risk management is identification of potential supply interruptions and devising contingency plans for them. This supply interruption can come from environmental disaster and can indirectly affect the firm. Therefore, gaining transparency and tracking a long supply chain to more than immediate suppliers is beneficial. One firm mentioned that due to tsunami in Japan there was a disruption in supply of semiconductors. The semiconductor company were 3rd or 4th tier supplier to them and if they had managed to get awareness on situation they would have been in better position. So there are suppliers that are in a critical chain of value chain and dependency on them is risky. The risk map and mitigation plan can be developed easier when there is transparency on situation and timely alerts. He explained that we **want a clear risk map that we click on supplier and it shows risks in red**. Currently, it is manual and not comprehensive. Surely, environment is not the only source of disruption and there can be political issues, financial problems and strike to cause disruptions. The point is to understand that how broad Big Data analytics should go to provide these insights.

If two company merge, their business position becomes important from risk management perspective and there should be business continuity alerts. One company was managing risk with following approach, getting information from RFI, RFQ and on site audits to create risk profile for suppliers and finding out their financial and business situation and availability of certifications in place. Weak points are identified through these profiles and mitigation actions are planned. They expected advanced analytics to provide transparency to shipments, quality feedbacks and experiences between company and for the data to be readily available when needed. In addition to material availability and disruptions, industries that face ethical or legal issues have risks regarding on issues such as food fraud, aviation legislation and labour force trafficking must track their long supply chains. Mapping a supply chain is not always the problem but the manual effort needed for it is cumbersome. One interesting point made during interviews was that connection of risk management to benchmarks in order to understand effects of risk on benchmarks which will be useful alongside running scenarios. In short, risk management is about early warning and alerts on factors that affect material availability. However, some companies were outsourcing risk management or it was not an issue for example in travel category, however it was a significant concern for others:

"Risk management is one topic we are discussing a lot and we need to kind of think what is the tool for risk management and also I don't know if it is possible to get that kind of future view from that point of view, also that can connect to benchmark information if everyone is starting to do something. Maybe early signs of some problems. Political situation, what is going to happen in the world but is it possible to get early. If there is happening something, we need to be really early warned and react before it comes to our face" (senior procurement manager case 5 food)

Table 7 Expectations from Big Data Analytics

Transparency	Automation	Prediction	Data Sources
• Central system providing transparency • Building big picture and filtering human's opinion • Transparency to all market related information • Transparency to search for different industry segments or technology segments or product globally and find supplier and supplier information available • Transparency to all the qualifications set, supplier audit, financial check, risk assessment, action plans, relationship plan, whatever comes with running the relationship and the business together	• One click getting all relevant information • Data collection through a tool for speeding up the process • Integration of big data from various departments • Data available and fetched automatically • Easier, more accurate way of identifying suppliers	• Optimising storage levels, forecasts and drawing conclusions • Use of distributions and regressions in decision making • Travel pattern and predicting holiday seasons • Forward looking from delivery capacity and cost perspective	• Academics • Literature of international organizations • Supplier market share, supplier businesses, pricing data • Recording of meetings • Price benchmarks • Competitors actions

Table 8 Expectations from Big Data Analytics (continued)

New KPI	Data representation	Others
• Variety on KPI other than on time delivery and quality • Measuring supplier flexibility	• Integration of multiple platform • One set of numbers, reports and future understanding • More technologically advanced than Excel • Visualizing through pictures not text	• Correct data at right time for right person • Data analytics for all components of TCO • Getting cost data per supplier, per equipment, per category • Extranet with suppliers for sharing information • Bringing stakeholders on board with transparency of analytics

4.4 Role of social media

Social media is growing source of data that is worthwhile investigating its usefulness, therefore interviewees were asked if they see any value in it. There were mixed opinions on the issue. Not all the companies understood opportunities and risks of social media but one mentioned that there are a lot of excitement to understand it better but they did not know how to use it. Additionally, it needs huge investment and change management and it should be done consistently to say whether it was success or not. There is no structured way of utilizing social media. Of course the perception of social media can be different industry wise. For instance in case of aviation company, they were mostly doing business to business (B2B) services that they were not in touch with second tier customer, so it was not a big thing for their procurement. Another point is following company's webpages and social media is time consuming and data might not be even relevant.

"I think there would be opportunities but companies are not yet using the opportunity and there is no good way currently to actually integrate multiple sources of information. It always requires one person to take the effort to go and take a look and in daily business many times you just don't have the time" (CPO case 2 engineering 2)

On the other hand case company 4 was curious about people's choice and what are the reasons behind their decisions. She mentioned that they need to understand customers' resentment or satisfaction and whether there is legitimate reason behind it or just anger. So it makes it risky to talk about products in social media and it needs moderation of these talks and any reaction that should be taken based on them. Another interviewee saw social media important because of increased demand for traceability and discussions around that topic in social media, so it is more of risk management to anticipate what happens to company before it is too late. Finally, one respondent found news valuable especially if it is political and can potentially link to company situations.

4.5 Risks of Big Data analytics and trusting Big Data analytics

Regarding risks, security and confidentiality of data was the most important concern of companies which puts a lot of focus on system access management. When the firm is operating within multiple countries then it becomes important what to share and what not to share. Furthermore, incorrect data poses risk since decisions made based on it, will be wrong. With volume attribute of Big Data, accessible data becomes huge and it is the issue of how to get the

data, how to ensure it is relevant and ensuring that there is compliance with regulation. In case of adoption of AI, it needs coaching to provide right answer and it was believed that people will stop doing something when they easily get insight from system. One interviewee was very concerned about how their data can be used by others and what kind of implications it will have in their position with supplier:

"How is our data potentially being used by others? Competitors and even non competitors. Third party analytics provider has local data, let's share supplier data with other companies. Ok, we would be interested to know what other companies are using but then at the same other companies will know what we are using and then ***they can take capacity out of our suppliers****.* ***Agreeing higher price levels with our suppliers which will then put pressure on us to get higher prices from same supplier then****" (SVP of global sourcing case 6 industrial machinery)*

When it comes to trusting data analytics, one person believed that they can be trusted as long as they are tested beforehand and they are not early adopters. Furthermore, management will trust data analytics more if there are improvements in quality of data. In decision making both politics and data can play a role and it is better to have data analytics rather than relying solely on politics. However, people are reluctant towards change and not all managers share the same view about data analytics:

"Some managers can state that data is not needed in environment where most of decisions are taken based on politics and based on experience of individual managers with varying interest. I believe this is one of the most destructive things that can happen" (category specialist case 3 aviation)

Almost all the companies considered data analytics as a tool to provide further insights and recommendation but none of them expected it to take actions or decisions. In decision making there should be a human factor and data analytics can act as an aid which is basis of decision making.

5 Discussion and conclusion

In this part, results are discussed in order to achieve answers for proposed research questions and limitations and further research opportunities are explained.

5.1 Answering research questions

Research Question 1: *What potential uses and needs companies are having in the area of procurement that can be supported with Big Data and advanced analytics?*

In order to investigate Big Data implications in procurement, supply base understanding was considered. The rationale behind that was to understand how complex a supply base can be and how firms identify their supply base to see if there are opportunities for Big Data to be useful in this area. Complexity of supply base differed in different categories that were chosen for interview and there were unique requirements for selecting a supplier. The most common practice among firms were using some reference whether through colleagues or consultancy firms to detect new suppliers. Some relied on industry fairs since participants had gone through a due diligence process that ensured they are following specific standards that was required in that industry. Public data is a potential source but reliability becomes important in that case.

There was no one size fit all practice among firms, however it was mentioned that **there is a need for a more efficient way of identifying supplier** to prevent missing opportunities. Big Data can help in mining public data to explore new suppliers but to ensure reliability, procurement solution providers can have important roles since suppliers might be their customer as well. It was also mentioned in interviews that in optimal case there would be **an extranet** that data is shared with suppliers through data. In that case procurement solution providers can play important role in creating the extranet since, they are able to compile data from different customers which comprise various industries. Therefore, they can provide information on available suppliers with higher level of reliability in case of specific requirements. Definitely, one should consider that it is highly reliant on consent of firms to provide their data or to agree on sharing it with other firms. Not all companies are willing to do it. One firm was concerned that revealing information of their suppliers can take away capacity of supplier and mitigate price negotiation position. Therefore, it is a matter of **change management and cost benefit analysis of such transparency**.

The second issue that is important regarding supply base is having view on n-tier suppliers which is directly linked to risk management. Risk management tries to ensure continuity of supply and avoiding disruptions. There can be ethical and legal considerations that pose risks for corporate image as well. While it was deemed unimportant in some categories, in overall it was important for firms to have transparency over their supply chain. Disruptions are not only caused by environmental disasters but there can be political, financial and social reasons behind them. The expectation from an advanced system is to **send timely alerts** regarding potential problems with critical suppliers for making contingency plans. Furthermore, transparency over suppliers helps in price negotiations.

Nevertheless, practicality is important in case of visibility over n-tier suppliers. While mapping a supply base might be easy, having actual visibility requires interaction with not only direct suppliers but 2nd or 3rd tier as well. Therefore, forming strategic relationships becomes important in achieving transparency, however it would be difficult to form such relationships with upstream suppliers. Willingness on sharing data should not be overlooked otherwise the available source would be public data which might not be accurate or reliable. Nonetheless, public data can reveal insights that are worth further investigation. Developing control towers mentioned by Mena et al (2014, p.94-95) seems a difficult task. Not only it requires cooperation in sharing data but connecting to supplier's systems requires aligning heterogeneous systems with different formats and technical specifications. Finally, it will require huge investment for creating such monitoring system.

The most common analytics available in companies is descriptive analytics. Even though firms show eagerness towards implementing Big Data, they do not completely know what kinds of analytics are available and what their capabilities are. In addition to n-tier suppliers monitoring and risk management, the main findings regarding expectations from Big Data were provision of transparency, automation, prediction, new data sources, new KPIs and new representation of data.

As Sanders (2014, p.131-147) mentioned order processing and automation as use cases, same result emerged in data regarding automation. It is not only limited to order processing but automatic data collection, integration of all departments' data and more accurate way in supplier identification.

Predictive edge proposed by Sanders (2014, p.131-147) was emphasized in interviews as more of forward looking system. This exhibits that companies see the need to move past descriptive

analytics. Sanders (2014, p.131-147) saw predictive analytics and scenario planning as solution to supply disruptions and risk management. Supply professionals also expected a risk management tool to give timely alerts and warnings. New KPIs other than quality and on time delivery should be developed. One proposed KPI was supplier flexibility that measures suppliers' resilience towards their customers. This finding is backed by Sanders (2016) that mentioned firms are dealing with measurement minutiae and they need new strategically aligned KPIs and the supplier resiliency score which was suggested by Sanders (2014, p.131-147). Additionally, he points out visualizing data in formats easy to understand. This claim is supported by firms' preference to have something more advanced than Excel and visualization through pictures.

Online literature and trade journals as sources of data which were mentioned by literature was verified based on interviews. In addition to that, **price benchmarks, competitors' situation and even recording of meetings** were potential sources of data. Potential sources of data can vary based on the category discussed and its requirements. Gandomi et al (2015) explained utilization of social media analytics in business. Results of this research displayed that companies have not completely thought about social media use case in procurement but for software category and food industry it was more important. Clearly there should be benefits to utilizing social media analytics especially in risk management. Stakeholder management is an important part in project management and social media can be a mean to interact with stakeholders. People can form political and social campaigns against companies or projects in different parts of the world that can lead to business continuity risks. Social media is a window to discussions that lead to forming opposing campaigns and gives the opportunity to firms to devise contingency plans or not even taking the risk.

Research Question 2: *What are the challenges and hurdles in implementing Big Data analytics in procurement?*

Most of the problems in implementing advanced analytics stem from state of internal data. Limberakis (2012) explained that most problems come from quality of internal data and efforts for cleansing and classifying data are mostly manual. Automation was suggested to improve internal data quality. Same problem appeared in companies as well. Some managers needed to collect data from different sources manually.

Internal data is a significant barrier in a firm's analytics journey. The quality issues raised by it, damages the trust needed to drive forward analytics journey. Even though firms are eager to

bring in Big Data capabilities, they were not fully trusting it because of quality issues and it was seen as something good to have in their decision making, not something of high importance. Moving towards Big Data analytics requires a cultural change in firm and acceptance of data driven decision making paradigm. Richey et al (2016) mentioned that Korean managers had emotional obstacle that guided them to base their decisions on past experience. This was not necessarily limited to Koreans. One of interviewees mentioned the existence of politics in decision making and how some managers are sceptics when it comes to data driven decision making. This is the mentality hindrance that Richey et al (2016) mentioned. Adopting predictive analytics and then prescriptive analytics needs a movement and management of resistance towards change. Therefore, it justifies the fact that Holsapple et al (2014) considers the movement (cultural change) as the first building block of business analytics. There are data ownership and security concerns that also stand in the way. Therefore, it is necessary to provide what can be achieved through Big Data to managers in order to gain their trust. One manager who was concerned about their data sharing could not see any proof of benefits outweighing disadvantages. Big Data initiatives have to bring stakeholders on board to realize the benefits since making cost benefit analysis is also challenging in view of Richey et al (2016).

Lack of awareness of Big Data opportunities might signal that firms do not trust analytics completely and not sure of their benefits. Nevertheless, quality issues of internal data exacerbates the situation. In fact the internal data impacts this research to find further use cases of Big Data since **internal data is a problem that needs to be fixed so that supply managers can envision further ahead and what more potential benefits can be gained**. One manager specifically mentioned that she wants the issue with internal data to be solved before anything else. Limberakis (2012) suggest, in order to make improvements, one firm should first start fixing problems regarding quality and integration and then move to more advanced features. In fact internal data and gaining trust of managers are main hindrance to Big Data application and they are interrelated. Also standardization of data is a huge effort that needs aid from 3rd part application providers provided that they have domain knowledge. Also it is a matter of ROI to convince top management.

A good way to show the situation is comparing results with Rafati & Poels (2015) view on data analytics in procurement. In the application layer there are various system that are not connected and coordinated. This was confirmed in the collected data of this research. There are system that are partially implemented around the world. Not all the functions work similar to

each other in each country. Such findings highlights the importance of having one version of truth (Richey et al, 2016) and improving value of Big Data by integration of data sources (Wang et al, 2015).

Data quality suffers because of various naming conventions and formats and data compilation is cumbersome due to dispersed system and sources. The deviation caused among internal data needs cleansing effort as well. In application layer it was seen that not all processes are in digital format or supported by system, this was seen in results form various Excel sheets and manual effort of managers to collect data and compiling it. In data layer, there are various system and standalone platforms that were not interconnected to have one set of number and one set of understanding which will lead to deviation and ultimately losing trust in data. Harmonizing internal data across functions alongside automation will change isolated parts of supply chain into an integrated one (Sanders, 2014, p.131-147)

Rafati & Poels (2015) suggest that the problem with application layer is lack of tools. However, firms are mostly using descriptive analytics. It might be true that there are not tools for prediction and deriving insights from data but it is the internal data state that is the inhibition to moving to next stage of analytics. In final layer or decision layer visualization techniques were suggested and companies expected the same, to have pictures over text and having something more advanced than Excel.

Finally, adopting Big Data analytics should be a calculated move. Snaders (2016) describes that companies do not know where to start and they need to have a systematic view so that all functions benefit from Big Data without increasing costs for other functions. Big Data is merely a tool for gaining competitive advantage. The hype and buzzword of Big Data should not become something similar to dot com bubble. Big Data itself do not bring added value. It requires fundamental changes in processes of data collection and storing to be effective and reliable.

5.2 Limitations and future research

This study was conducted to elicit needs and expectations of supply and purchasing professionals from Big Data analytics. The focus of this study was placed on procurement function and how it can benefit from Big Data. It should be noted that interviewees of this study were viewing this topic with their definition of Big Data and how familiar they were with it. Additionally, one should consider that Big Data is an evolving concept that is under

investigation and there is no concrete definition for it in minds of people. Challenges regarding internal data were identified, it was not only limited to internal data of procurement but other departments that procurement relies on them as well. However, the situation with other functions and how they handle their systems were not investigated. Therefore, it is not clear whether the high quality data needed for procurement can be provided and what issues other departments are facing to provide.

This study was a qualitative explorative study trying to understand needs of supply and purchasing professionals regarding Big Data. Consequently, results of this study touch the surface of topic and give a holistic view on what the expectations from Big Data are. Therefore, investigating results in detail can be topic for future research. For example, it would be worthwhile to dive deeper into benefits of Big Data in risk management. Firms are not fully aware of Big Data capabilities, future research can focus on co-creation of capabilities with firms and service providers. Service providers can propose available capabilities of Big Data to companies so that companies can also add value to that based on their own views and new ideas can be generated by both parties.

This study was not limited to Finnish companies and international companies were included as well. Interviewing not just CPOs but category managers proved to be useful since it provided various perspective on issue which is valuable in designing future researches. It was not possible to have multiple interviews for all the case companies, however future research can benefit from multiple interviews per case company. The firms studied in this research were end users of Big Data analytics. Future research can include service providers and especially consultancy firms that are sources of data for end users.

6 References

Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A. and Buyya, R., 2015. Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79*, pp.3-15.

Baxter, P. and Jack, S., 2008. Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report*, *13*(4), pp.544-559.

Chen, H., Chiang, R.H. and Storey, V.C., 2012. Business intelligence and analytics: from big data to big impact. *MIS quarterly*, pp.1165-1188.

Chen, D.Q., Preston, D.S. and Swink, M., 2015. How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems*, *32*(4), pp.4-39.

Chithur, D. 2014. Driving Strategic Sourcing Effectively with Supply Market Intelligence. Tata Consultancy Services (TCS).

Chowdhary, P., Ettl, M., Dhurandhar, A., Ghosh, S., Maniachari, G., Graves, B., Schaefer, B. and Tang, Y., 2011, October. Managing procurement spend using advanced compliance analytics. In *e-Business Engineering (ICEBE), 2011 IEEE 8th International Conference on* pp. 139-144. IEEE.

Demchenko, Y., Grosso, P., De Laat, C. and Membrey, P., 2013, May. Addressing big data issues in scientific data infrastructure. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* pp. 48-55. IEEE.

Fan, Y., Heilig, L. and Voß, S., 2015, August. Supply chain risk management in the era of big data. In *International Conference of Design, User Experience, and Usability*, pp. 283-294. Springer, Cham.

Gandomi, A. and Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), pp.137-144.

Gao, J., Koronios, A. and Selle, S., 2015. Towards a process view on critical success factors in big data analytics projects.

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S.F., Childe, S.J., Hazen, B. and Akter, S., 2017. Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, *70*, pp.308-317.

Hazen, B.T., Boone, C.A., Ezell, J.D. and Jones-Farmer, L.A., 2014. Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, pp.72-80.

Hofmann, E., 2017. Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. *International Journal of Production Research*, *55*(17), pp.5108-5126.

Holsapple, C., Lee-Post, A. and Pakath, R., 2014. A unified foundation for business analytics. *Decision Support Systems*, *64*, pp.130-141.

Karthik, V. and Gupta, N., PROCUREMENT ANALYTICS EFFECTS ON ORGANISATION. In *Sustainable Development* by Dubey A.D and Thomas J. Delhi, India: Research India Publications, pp.13-27.

Keith, B., Vitasek, K., Manrodt, K. and Kling, J., 2015. *Strategic Sourcing in the New Economy: Harnessing the Potential of Sourcing Business Models for Modern Procurement*. Springer, pp.233-234.

Kraljic, P., 1983. Purchasing must become supply management. *Harvard business review*, *61*(5), pp.109-117.

Koutroumpis, P. and Leiponen, A., 2013, October. Understanding the value of (big) data. In *Proceedings of 2013 IEEE international conference on big data* pp. 38-42.

Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, *6*(70).

Limberakis, C.G., 2012. Spend Analysis: LESSONS from. *Supply Chain Management Review*, *16*(2), pp.10-12.

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.

Mena, C., Christopher, M. and Van Hoek, R., 2014. *Leading Procurement Strategy: Driving Value Through the Supply Chain*. Kogan Page Publishers, pp.94-95.

Pavlou, P.A. and El Sawy, O.A., 2011. Understanding the elusive black box of dynamic capabilities. *Decision sciences*, *42*(1), pp.239-273.

Pavlou, P.A. and El Sawy, O.A., 2011. Understanding the elusive black box of dynamic capabilities. *Decision sciences*, *42*(1), pp.239-273.

Press, G. (2014) *12 Big Data Definitions: What's Yours?* [online]. Available from: https://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#72b853d813ae (Accessed 4 April 2018).

Rafati, L. and Poels, G., 2015, February. Towards model-based strategic sourcing. In *International Workshop on Global Sourcing of Information Technology and Business Processes* pp. 29-51. Springer, Cham.

Richey Jr, R.G., Morgan, T.R., Lindsey-Hall, K. and Adams, F.G., 2016. A global exploration of big data in the supply chain. *International Journal of Physical Distribution & Logistics Management*, *46*(8), pp.710-739.

Sallam, R.L., Richardson, J., Hagerty, J. and Hostmann, B., 2011. Magic quadrant for business intelligence platforms. *Gartner Group, Stamford, CT*.

Sanders, N.R., 2014. *Big Data Driven Supply Chain Management: A Framework for Implementing Analytics and Turning Information into Intelligence*. Pearson Education, pp.131-147.

Sanders, N.R., 2016. How to use big data to drive your supply chain. *California Management Review*, *58*(3), pp.26-48.

TAN, H. and LEE, W.L., 2015. Evaluation and Improvement of Procurement Process with Data Analytics. *International Journal of Advanced Computer Science and Applications*, *6*(8), p.70.

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

Schoenherr, T. and Speier-Pero, C., 2015. Data science, predictive analytics, and big data in supply chain management: Current state and future potential. *Journal of Business Logistics*, *36*(1), pp.120-132.

Teece, D.J., Pisano, G. and Shuen, A., 1997. Dynamic capabilities and strategic management. *Strategic management journal*, *18*(7), pp.509-533.

Vossen, G., 2014. Big data as the new enabler in business and other intelligence. *Vietnam Journal of Computer Science*, *1*(1), pp.3-14.

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

Wang, G., Gunasekaran, A., Ngai, E.W. 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*, pp.98-110.

Yin, R.K., 2015. *Qualitative Research from Start to Finish*. New York, NY: Guilford Publications