CHALMERS
UNIVERSITY OF TECHNOLOGY

Possibilities in using big data to improve supply chain efficiency

A study in the inbound logistics of an automotive company

Master of Science Thesis
in the Supply Chain Management master's Programme

ITZEL BAÑUELOS REYES
GEORGIOS RODITIS

Department of Technology Management and Economics
Division of logistics and transportation
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2015
Report no E2015:084

MASTER'S THESIS E2015:084

Possibilities in using big data to improve supply chain efficiency

A study in the inbound logistics of an automotive company

ITZEL BAÑUELOS REYES
GEORGIOS RODITIS

Examiner: Per Olof Arnäs

Department of Technology Management and Economics
Division of logistics and transportation
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden, 2015

Possibilities in using big data to improve supply chain
A study in the inbound logistics of an automotive company

ITZEL BAÑUELOS REYS
GEORGIOS RODITIS

Master's Thesis E2015:084

Department of Technology Management and Economics
Division of logistics and Transportation
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Telephone + 46 (0)31-772 1000

Chalmers Reproservice
Göteborg, Sweden 2015

Abstract

The purpose of this thesis project was to investigate the possibilities of using big data for the material transport process (inbound logistics) of an automotive company. The study was performed in two stages: An exploratory phase and the development of a proof of concept. The first stage allowed the understanding of the material transport process at a strategic, tactical and operational level, as well as the identification of fourteen big data opportunities that support the drivers of the focal company. In the second stage, the CRISP-DM methodology was applied to develop a proof of concept that would transform the material forecast into a transport forecast. It was identified, that the development and implementation of this proof of concept would provide benefits in a tactical level, such as: improved volume optimization, better price negotiation with carriers, improved management of human resources in terminals and better cost control, among others. In order to get these benefits, the focal company may overcome three main gaps: the accuracy and reliability of information in the system, the level of detail and the quality of the data. It was concluded that the use of big data will have a significant positive impact for the focal company.

Key words: Inbound Logistics, Transport, Big Data, Efficiency, Barriers, Benefits, Automotive, Visibility, Forecast, Supplier, Carrier.

Preface

This thesis was conducted at Chalmers University of Technology in Gothenburg, Sweden, during the spring of 2015 as a final part of the Master Program of Supply Chain Management. The thesis has been a cooperation with an automotive company with headquarters in Sweden.

Firstly, we would like to thank Mathias Humlefjäll, for giving us the opportunity to do this job; his time, knowledge, advice and the freedom he has given us to think and decide were key. Thanks for your friendship and support!

We would also like to thank Per Olof Arnäs, his guidance and experience were key in the decisive moments; while his feedback as a professor and as a mentor was invaluable. Thank you PO!

Itzel and Georgios

I would like to dedicate this job to my family: Christian, for all the time he has spent on it (reading, helping and even taking care of everything so I could study or work) and for all the joy and love he brings into our lives; my parents, Yolanda and Hector, for being my life inspiration, always encouraging me to work harder and achieve more; my sisters, Sara and Naye, for their unconditional love and support, always cheering me up and believing in me; my parents-in-law, Anki and Erland, my brothers-in-law Paco and Kakalos and my nieces, Andrea and Elena; my friends, Abner, Silvana and Reynaldo; my co-writer, and friend, Georgios, for joining me in this venture: it was a pleasure to work with you. And at the end, I would like to thank specially to the most special person in my life, Oscar, who joined me since the beginning of this journey and makes every day a better day: thanks for existing.

Itzel

I would like to dedicate this work to all my beloved ones beginning from my family in Greece: my parents Anastasia and Alexandros for their psychological support throughout the process and my sister Stella for her efforts to cheer me up during the bad times. In addition, I should mention some of my close friends that also supported me and are always available to help me in anything: to Abner Braga my dear friend and fellow student from Brazil, to Konstantinos Sofos and Irene Zira, my friend Giannis Kelaris and of course my friend, fellow student and co-writer Itzel for all her drive and dedication to this project in both good and bad times. Last but not least, since she is accustomed to get on stage and close the show, I would like to dedicate this work to my girlfriend Dorina for her patience, love and concern, she acted like beacon of light for this journey. Thank you all my dear people.

Georgios

Table of Contents

List of figures

List of tables

List of Abbreviations

3PL	Third party logistics
BOM	Bill of materials
CRISP-DM	Cross industry standard for data mining
ETA	Estimated time of arrival
ETD	Estimated time of departure
FTL	Full truck load
LTL	Less than a truck load
POA	Proof of arrival
POC	Proof of collection
POD	Proof of delivery
TB	Transport Booking
TBR	Transport Booking Request
THU	Transport Handling Unit
TS	Transport status

1 Introduction

Globalization, technological trends and growing dynamism in sourcing strategies are some of the challenges that companies have to face nowadays: they represent an opportunity to improve performance, but also an increasing demand and pressure on its supply chains, as complexity and fragmentation of operations have increased (Nabben, 2014; DHL, 2015).

According to Harrington and Smith, (2013), for the specific case of the automotive industry, four trends are shaping the sector: global growth and emerging markets, products closer to customer needs, relentless cost pressure and investment in mega plants in overseas locations like China or Mexico. The combination of these is increasing the pressure to improve responsiveness, reliability and resilience while decreasing costs (Harrington & Smith, 2013; Christopher, 2011). For the logistics functions of the supply chain, this will translate into efforts to decrease transport costs, inventories and administrative costs, while maintining or improving flexibility, delivery reliability, availability and capacity utilization (DHL, 2015).

The technological advancements in recent years have allowed companies to adopt better tools to face challenges. A lot of attention has been paid in recent years to the term ***'big data'***, both from academia and businesses (Waller & Fawcett, 2013). The increase in the use of enterprise systems, mobile devices and RFID tags, as well as the explosion in the use of social networks, have allowed companies to collect large amounts of information (Accenture, 2014) deriving in the term ***'big data':*** "big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the structures of your database architectures" (Dumbill, 2012, p. 3).

Big data is currently being used in a wide variety of fields, from marketing intelligence to crime prediction and certainly, it represents a big potential for Supply Chain (Waller & Fawcett, 2013). According to a research done by Accenture (2014) on 1014 companies 97% of executives have an understanding of how big data can support their supply chain, but only 17% have been able to implement it. Data itself represents an opportunity to companies, as it can help to make better decisions in an operational, tactical and strategic level (Schroeck, et al., 2012), with direct impact in profitability (Waller & Fawcett, 2013).

The logistics network of the focal company consists of several thousands of senders and about 1,000 receivers. With the aim of connecting them, more than a thousand carriers are contracted to perform more than ten millions shipments in a year. As a consequence, the focal company is both a big generator and a big user of information. Large amounts of data is created every minute, while at the same time, different kinds of information are necessary to perform adequate analysis and make decisions in the short and the long term. The possibility of analyzing and processing this data, as well as integrating other sources of information in its processes, through the use of big data and its tools, analytics or methodologies may represent an opportunity for the focal company.

1.1 Background

The focal company is one of the leaders in the automotive manufacturing, with headquarters in Europe. It has production facilities in Asia, America and Europe, and sales all over the world. The company has a specific logistics department, in charge of providing all the inbound and outbound logistics services, i.e. organizing the activities related to the movement of materials from material suppliers to the assembly plants and the finished products to customers. The main objective of the department is to design, manage and optimize the supply chain.

The amount of logistics suppliers, daily transactions as well as the number of systems or applications used within the material transport process, makes it difficult for the focal company to reach operational efficiency in five different domains: network design, network configuration, planning and optimization, execution and analytic insight (Figure 1). Within its internal data repositories or in other external sources, there is data that can potentially support the company to make quicker and more accurate decisions, as well as performing deeper analysis of the operations.

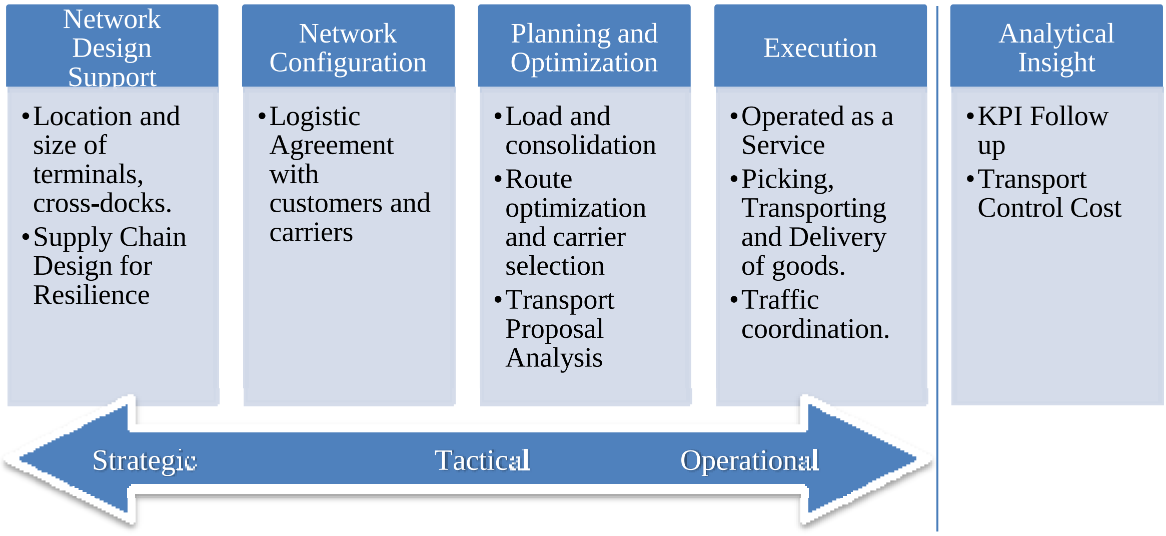


Figure 1 – Five domains in the transport material process

Previous studies have been able to identify the benefits of big data for the supply chain or logistics. For example, a survey done by Accenture (2014), showed that 41% of the companies that implement big data analytics achieved faster and more effective reaction time to supply chain issues, 36% increased supply chain efficiency by 10% or more, 33% were able to optimize inventory and asset productivity and 28% got better customer and supplier relationship (Figure 2).

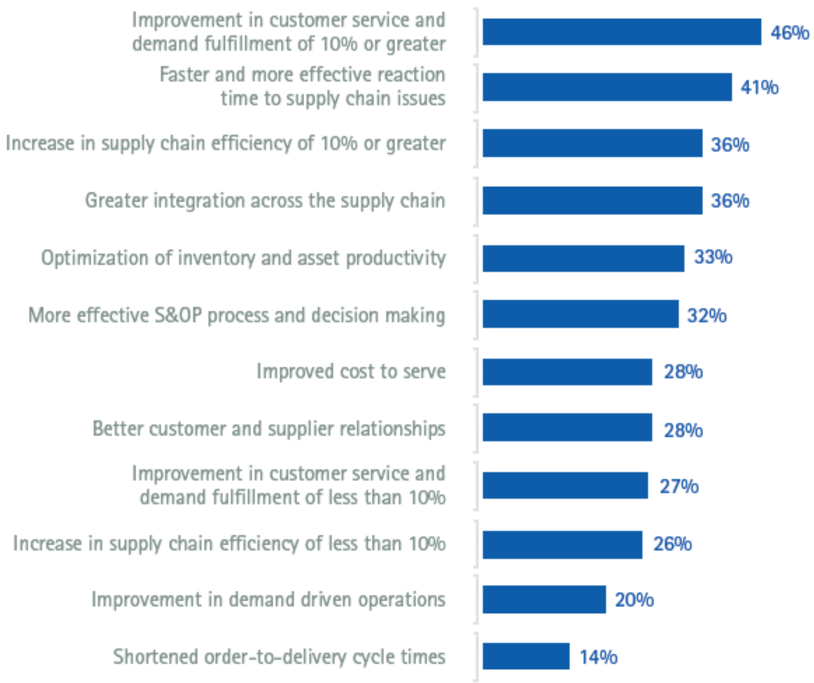


Figure 2 – Results companies have achieved by using big data analytics (Accenture, 2014).

DHL has named big data one of the most important current trends (Bubner, et al., 2014) and, through a cooperation with T-Systems (Jeske, et al., 2013) was also able to identify three main areas where big data can provide value to businesses: operational efficiency, as data can increase transparency, optimize the use of resources and improve quality and performance; customer experience, as it can benefit interaction with customers and increase customer loyalty; and data capitalization, as new revenue stream can be created when seeing data as a product.

1.2 Purpose

After reviewing literature, it was identified that big data is a current trend that is supporting many companies to reach operational efficiency within their Supply Chain. The focal company aims at investigating what the opportunities of big data are.

This thesis will focus on exploring the utility of big data within the material transport process in terms of complexity, benefits, enablers and barriers.

1.3 Research questions

Taking as a reference the results that other companies are achieving through the use of big data, the purpose of this study is to answer the following research questions.

Research Question 1: How can the material transport process benefit from big data on a strategic, tactical and operational level?

Research Question 2: What are the benefits, costs and barriers in implementing big data for the material transport process?

1.4 Scope and limitations

The scope of this Master Thesis is mainly focused on the operational efficiency of the supply chain as a method to improve performance in terms of responsiveness and cost: the environmental and social aspects related to logistics operations have not been considered in detail when analyzing the logistics operations of the focal company, as currently they are not the main challenges of the logistics department.

The benefits of the all the big data opportunities were identified. Nevertheless, due to time limitations, only the opportunity developed into a proof of concept was further analyzed into barriers stage.

Due to the complexity and advanced technical level of big data, the evaluation of the costs of implementing big data was done in terms of complexity, not economic.

1.5 Outline

Methodology	This chapter describes how the research was conducted: what methods were used to study the literature and to conduct the interviews and surveys needed for this research. The methodology CRISP-DM, used to develop the proof of concept, will be presented.
Frame of reference	In this chapter, the theoretical background for Logistics will be described. Concepts such as efficiency, responsiveness, visibility will be discussed. Additionally, the definitions of big data, big data analytics and data science will be presented, together with the benefits of big data identified by different researches.
Empirical Findings	This chapter will present information about the inbound logistics process and how the logistics functions are organized in five domains. The material order process and packaging instructions process will be shortly described due to its relevancy for the development of the proof of concept.
Pre-Analysis	A pre-analysis stage was conducted with the aim of identifying a set big data opportunities that could provide benefits to the focal company. The purpose is that one of this ideas will be further developed into a proof of concept.
Proof of concept	This chapter presents the proof of concept developed: from the evaluation of the fourteen big data opportunities and the selection of one of them, until the model development and assessment. To ease the understanding of the development and findings, the information is presented following the phases of the selected methodology, CRISP-DM.
Analysis and Discussion	In this chapter, the results of the proof of concept and its future possibilities, as well as the fourteen big data opportunities will be further discussed in terms of benefits, enablers and barriers. The results will be contextualized in terms of the relevant theoretical framework.
Conclusions and Recommendations	Finally, the study will be summarized in this chapter. Conclusions from the big data opportunities and the proof of concept will be presented, together with recommendations for the focal company and future academic research.

2 Methodology

The type of investigation is followed is a case study as it is defined by (Yin, 1994, p. 13) “a case study is an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident”. In addition to this definition the author elaborates on the technical aspect: “The case study inquiry copes with the technically distinctive situation in which there will be many more variables of interest than data points and as one result relies on multiple sources of evidence, with data needing to converge in a triangulating fashion, and as another result benefits from the prior development of theoretical propositions to guide data collection and analysis”.

The research questions formulated in the previous chapter belong to the “what & how” types of questions, (Yin, 1994) used in exploratory case studies. Thus, the type of the scientific research that was followed in this research is the exploratory case study.

The case study was built on four types of research: a literature study, interviews, a survey and a proof of concept.

2.1 Literature Study

Booth et al. (2008) describe the source searching methodology that was used in order to conduct the literature study. This study was conducted in order to build a theoretical framework on big data in the logistics field. Thus, the theoretical framework was built on data collected through secondary sources like books, journals, databases available through the following methods:

- Search by keywords in the Chalmers Library computerized resources. Keywords: “big data”, “Data science”, “Logistics”, “Supply Chain Management” etc.
- Search by keywords in Scopus, the online searching tool of Chalmers Library
- Search following references from experts
- Search by keywords in Google scholar
- Snowball effect search, to use references of sources in order to access other sources.

The screening process described by Booth et al. (2008) includes prioritizing the titles and abstracts of the sources by relevance to the subject. All sources and the search criteria were documented and referred according to Harvard denotation as presented in the reference list.

2.2 Interviews

Kothari in his work (2004) describes the methods of data collection that are commonly used in research methodology and specifically the interview method. The interview method involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses. Two types of interviews are mainly used: Personal interviews and Telephone interviews. For the purpose of this study both types of interviews type was used.

Personal interview can be conducted in the form of direct personal investigation or in the form of indirect oral investigation (Kothari, 2004). In the case of direct personal investigation which was followed in this report, the interviewer has to collect the information personally form the sources concerned since this form is designed for an intensive investigation on specific matters. According to the author, telephone interviews consists of contacting respondents via telephone to collect the information needed, but for this study some of the interviews with respondents from other places of the world that could not be conducted personally, were done through the use of Lync, which is a software available in the market for teleconferences. During these interviews there was no face to face communication but sharing of screens and presentations took place.

The personal interviews were conducted in an unstructured way in order to maintain the flexibility of approach to questioning (Kothari, 2004). Unstructured interviews do not follow a system of pre-determined questions and standardized techniques of recording information. A greater freedom was allowed in these interviews regarding the sequence of the questions and the use of supplementary questions at points were clarifications or deeper knowledge was important to obtain. The unstructured type of interview is the central technique of collecting information in case of exploratory research studies such as this.

2.3 Survey

According to Kothari (2004) there are two types of surveys commonly used in research methodology: census survey and sample survey. All items in any field of inquiry constitute a “population”; a complete enumeration of all items in the “population” of the inquiry is called a census survey. In most cases this method is practically beyond the reach of ordinary researchers, governments are the main implementers of this method. Thus, for the purpose of this study the type of the sample survey was followed. When studies are undertaken in practical life, considerations of time and cost lead to a selection of only a few items of the “population” as respondents. The selected respondents constitute the “sample” and the study is called “sample survey”.

There are different types of “sampling” in a sample survey (Kothari, 2004). The sampling type that was used for the purpose of this thesis is the “non-probability sampling”. This type of sampling procedure does not afford any basis for estimating the probability for each item of the population to be included in the sample and is also called deliberate sampling, purposive sampling and judgment sampling. The items for the sample were deliberately selected for this report following this type of sampling. The degree of bias danger in this kind of sampling is high, however according to the author, in small inquiries and researches by individuals, this type of sampling can be adopted because of the relative advantage of time and cost inherent in it.

The form of the data collection method that was used under the survey is the questionnaire. A questionnaire consists of a number of questions printed or typed in a definite order on a form or set of forms (Kothari, 2004). The general form of the questionnaire used was the structured form. Structured questionnaires are those in which there are definite, concrete and pre-determined questions that are presented with exactly the same wording and in the same order to all respondents of the survey. The forms of the questions used in the questionnaire were closed (picking one defined answer) but there was also a level of freedom given to the respondent if they wanted to type comments freely.

The questionnaire was sent by e-mail to 23 recipients that constitute the survey sample. The total respondents of the survey were 13 corresponding to 57% of the sample. The sample was selected from several departments including: controlling, logistics cost analysis, tactical & operational optimization, purchasing, operations and strategy. Members of the sample were situated in different regions including: Europe, Asia and America.

The objective of the questionnaire was to identify benefits derived from a possible implementation of the suggested solutions to the problems identified in the previous steps of the process. In order to be able to quantify and use effectively the results of the survey an average benefit calculation was conducted.

2.4 Proof of concept

The proof of concept was done following the CRISP-DM methodology, a model specifically developed for analytics and data mining (Chapman, et al., 1999) that has been successfully applied in the academia and the industry (Kurgan & Musilek, 2006). The methodology consists of six phases (Figure 3), and includes feedback loops and iterations, which is beneficial for data mining projects (Kurgan & Musilek, 2006).

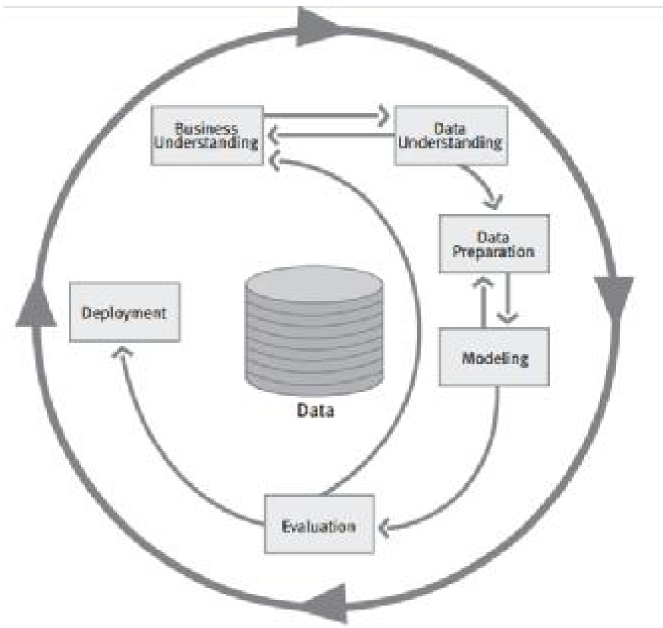


Figure 3 – Phases of the CRISP-DM reference model. (Chapman, et al., 1999)

Only the first five phases of the methodology were applied for the proof of concept developed in this study: the first four of them, are covered in chapter 5, proof of concept; the fifth phase, evaluation is covered in chapter 6, analysis; while the sixth phase, the deployment, which refers to the implementation, monitoring and maintenance, was not covered for being out of the of the scope of this thesis.

The complexity for the focal company was evaluated considering the number of activities that would take place in order to develop the project: data/service acquisition, data mining, matching/comparison of two or more data sources, analytical or info processing required, organizational impact and software development required. Nevertheless, due to the nature of this study, those opportunities that required a software development were ranked as 5 due to the high level of time and programming skills that would be required.

The big data index was calculated based on a methodology developed during this study, where the characteristics of big data, volume, velocity and variety were evaluated in a scale from 1 to 3.

2.5 Research quality of qualitative research

According to Riege (2003) there are different techniques to enhance the quality of a case study, based on four aspects widely used in social research: construct validity, internal validity, external validity and reliability. In this section, a description of each of the aspects is given, as well as which techniques were used in this study to ensure trustworthiness.

2.5.1 Construct validity

Construct validity refers to the term confirmability in qualitative research and describes the neutrality and objectivity of the study, assessing if the interpretation of the data has been performed in a logical and unbiased way (Riege, 2003). Key questions on this subject are how complete is the information gathered by the researchers and if the data is available for reanalysis by others (Miles and Huberman, 1994, cited in Riege, 2003, p.81).

The techniques used to construct validity are: using of multiple sources of evidence, creation of a chain of evidence as well as having key informants review the reports (Riege, 2003). In order to construct validity, sources taken from literature and internal sources of the company were used; the chain of evidence was secured by keeping notes of the interviews and the surveys which were reviewed by supervisors both from the university in order to secure the academic validity as well as from the focal company in order to secure the compliance of the report to the desirable results.

2.5.2 Internal validity

Internal validity refers to the term credibility in qualitative research and the aim is to prove that the reality corresponds to the researcher's interpretation (Riege, 2003). Key questions on this subject are how complete the descriptions are, and how coherent are the findings (Riege, 2003).

The techniques used to increase internal validity are the use of illustration and diagrams for further explanations, as well as cross-checking of the results (Riege, 2003). During this study, figures and tables were used and cross-checked with supervisors and key interviewees in order to assure the appropriate understanding of the processes described.

2.5.3 External validity

External validity refers to the term transferability in qualitative research and the aim is to achieve analytical generalization (Riege, 2003). Key questions on this subject are if the descriptions are long enough to assess transferability and if the results are congruent to prior theory (Riege, 2003).

One of the techniques used to increase external validity is comparing evidence with external literature (Riege, 2003). During this study, the results of this report have been compared to surveys conducted by other consulting firms and authors presented in the theoretical framework.

2.5.4 Reliability

Reliability refers to dependability in qualitative research and the aim is to stability and consistency achieve analytical generalization (Riege, 2003). Key questions on this subject are if the research questions are clear and if the research has been conducted carefully (Riege, 2003).

The techniques used to increase reliability are using multiple researchers, recording data, assuring parallelism of findings and using peer review (Riege, 2003). During this study, two researchers performed the study, which was supervised by external actors and was continuously reviewed by the researchers. The data collected from the interviews has been kept in the form of notes from the researchers; while the results of the survey have been kept in digital form.

3 Theoretical framework

3.1 Supply chain management and logistics theory

3.1.1 Supply chain management definition

Handfield and Nichols (2002, p. 8)developed the following definition of a supply chain: "The supply chain encompasses all organizations and activities associated with the flow and transformation of goods from raw materials, through to the end user, as well as the associated information and monetary flows."

The Council of Supply Chain Professionals (CSCMP, 2013) in their glossary of terms define Supply chain Management as:

> "Supply Chain Management encompasses the planning and management of all activities involved in sourcing and procurement, conversion, and all logistics management activities. Importantly, it also includes coordination and collaboration with channel partners, which can be suppliers, intermediaries, third-party service providers, and customers. In essence, supply chain management integrates supply and demand management within and across companies. Supply Chain Management is an integrating function with primary responsibility for linking major business functions and business processes within and across companies into a cohesive and high-performing business model. It includes all of the logistics management activities noted above, as well as manufacturing operations, and it drives coordination of processes and activities with and across marketing, sales, product design, finance and information technology." (CSCMP, 2013, p. 187) .

Chopra and Meindl (2013) give plenty of examples of companies such as Dell, Amazon, Wall-Mart, Seven Eleven and others that recognized the importance of applying supply chain management strategies to increase the profitability of theirs firms. In contrast, the authors provide with examples of companies like Webvan, Kozmo and Borders that failed to identify this importance and gradually failed to adapt to modern business environment. Supply chain management is necessary in a modern company in order to manage the supply chain and adapt it to changing technologies and customer expectations.

3.1.2 Logistics definition

Gleissner and Femerling (2013, p. 4) use the definition of The European Committee of Standardization CEN in order to define Logistics as: "The planning, execution and control of the movement and placement of people and/or goods and of the supporting activities related to such a movement and placement, within a system organized to achieve specific objectives".

The central tasks of logistics were graphically depicted in the form of the seven R's shown in figure 4 (Gleissner & Femerling, 2013). The seven R's constitute of the right elements respectively for products, quantity, quality, place, time, costs and customers.

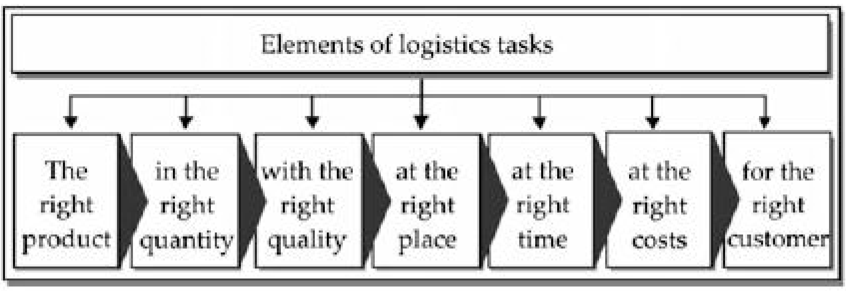


Figure 4 - Elements of Logistics Tasks (Gleissner & Femerling, 2013)

Logistics in general has been a field that has been growing for several years and will continue growing for the next years to come (Cheong, 2003). The Logistics field constitutes a very important part of every economy worldwide since for example the costs associated with logistics activities average about the 12% of the World GDP. Globalization was a major trend behind the growth of the logistics sector in the form of the development of the Third Party Logistics (3PLs) companies that are the main conductors of logistical activities since companies focus more and more in their core competencies and outsource the rest.

According to other Ansari and Modaress (2010), there are some key drivers at the side of the shippers that lead to the recent growth of the 3PL industry globally. There is a significant pressure to further improve operational efficiency in their operations, to improve the quality levels, to enhance flexibility, to secure short term profitability as well as to maintain a very high level of customer service. These drivers have forced a large amount of shippers to outsource several parts of their supply chain activities to 3PL companies.

3.1.3 Supply Chain Management Strategy & Logistics

There are several authors that have put effort in illustrating the connection between supply chain management and logistics. For the purpose of this thesis research, the theoretical framework that was used to set the foundations for the analysis was the one created in the book by Chopra and Meindl (2013). In their book the authors build a framework that illustrates how supply chain management strategy derives from the overall competitive strategy of a company and how it affects the performance of the company.

For the formulation of their framework the authors identified the two dimensions of supply chain performance of a firm: responsiveness and cost (Chopra & Meindl, 2013). Supply chain responsiveness is defined as the ability of the supply chain to respond to wide range of demanded quantities, meet short lead times, handle a large variety of products, build highly innovative products, meet a high service level and handle supply uncertainty. In contrast, the cost dimension is defined by the term supply chain efficiency which is the inverse of the cost of making and delivering a product to the customer, thus an increase in cost lowers efficiency.

There is a trade-off between supply chain responsiveness and efficiency: every strategic choice to increase responsiveness leads to additional costs that lower efficiency (Chopra & Meindl, 2013). Thus it is left to the companies' decision makers to find the best combination of the two dimensions of the supply chain performance. The differences that derive in the supply chains after they are characterized as efficient or responsive the authors is shown in figure 5.

	Efficient Supply Chains	*Responsive Supply Chains*
Primary goal	Supply demand at the lowest cost	Respond quickly to demand
Product design strategy	Maximize performance at a minimum product cost	Create *modularity* to allow postponement of product differentiation
Pricing strategy	Lower margins because price is a prime customer driver	Higher margins because price is not a prime customer driver
Manufacturing strategy	Lower costs through high utilization	Maintain capacity flexibility to buffer against demand/ supply uncertainty
Inventory strategy	Minimize inventory to lower cost	Maintain *buffer inventory* to deal with demand/supply uncertainty
Lead time strategy	Reduce, but not at the expense of costs	Reduce aggressively, even if the costs are significant
Supplier strategy	Select based on cost and quality	Select based on speed, flexibility, reliability, and quality

Figure 5 - Comparison of Efficient and Responsive Supply Chains (Chopra & Meindl, 2013, p. 42).

The framework that derives from the translation of the competitive strategy of the company to its supply chain strategy presented by Chopra and Meindl (2013) illustrates the two dimensions of supply chain performance in relation to the three logistics parameters and the three cross-functional parameters of the supply chain strategy, see figure 6.

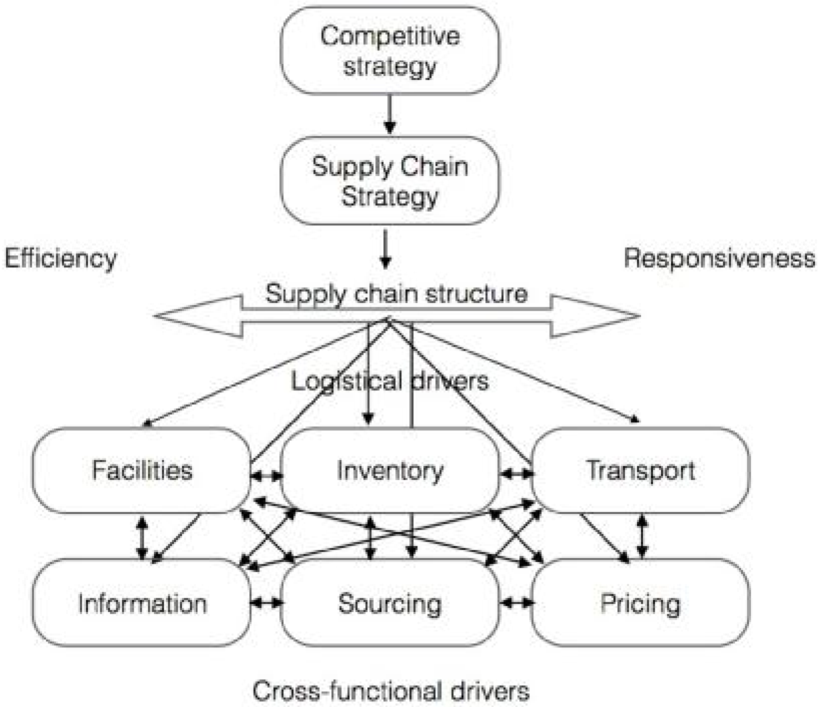


Figure 6 - Supply chain decision making drivers (Chopra & Meindl, 2013).

The three drivers of logistics depicted in the framework are: facilities, inventory and transport and the three cross-functional drivers are: information, sourcing and pricing (Chopra & Meindl, 2013).

The driver of facilities consists of the actual physical locations in the supply chain network where product is stored, assembled, or fabricated. The two major types of facilities are divided into production sites and storage sites. Decisions regarding the role, location, capacity, and flexibility of facilities have a significant impact on the supply chain's performance (Chopra & Meindl, 2013).

The inventory driver consists of all raw materials, work in process, and finished goods within a supply chain. The inventory belonging to a firm is reported under assets. Changing inventory policies can dramatically alter the supply chain's efficiency and responsiveness. Some of the most important metrics concerning inventory are the average inventory metric which is the average amount of inventory carried and the fill rate which is the fraction of orders/demand that were met on time from inventory (Chopra & Meindl, 2013).

Transportation driver is consisted of the movement of inventory from point to point in the supply chain. It can take the form of many combinations of modes and routes, each with its own performance characteristics. Transportation choices have a large impact on supply chain responsiveness and efficiency. Outbound transportation costs of shipping to the customer are typically included in selling, general, and administrative expense, while inbound transportation costs are typically included in the cost of goods sold (Chopra & Meindl, 2013).

Information consists of data and analysis concerning facilities, inventory, transportation, costs, prices, and customers throughout the supply chain. Information is probably one of the biggest driver of performance in the supply chain according to the authors because it directly affects each of the other drivers. Information presents management with the opportunity to make supply chains both more responsive and more efficient. As it argued by the authors, with significant investment in information technology solutions, the visibility and coordination of decisions in the supply chain increases, thus securing an essential prerequisite for the company's profitability (Chopra & Meindl, 2013).

Sourcing is the selection of the party that will perform a particular supply chain activity such as production, storage, transportation, or the management of information. At the strategic level, these decisions determine what functions a firm performs and what functions the firm outsources. In addition, the other two strategic decisions that are supported through the sourcing driver is the selection of the suppliers and the procurement. The criteria for the suppliers selections are set as well as the relationships between the company and its suppliers ensuring that the transaction costs are low (Chopra & Meindl, 2013).

Pricing determines how much a firm will charge for the goods and services that it delivers in the supply chain. Pricing affects the behavior of the buyer of the good or service, thus affecting supply chain performance. Differential pricing provides responsiveness to customers that value it and low cost to customers that do not value responsiveness as much. Any change in pricing impacts revenues directly but could also affect costs based on the impact of this change on the other drivers (Chopra & Meindl, 2013).

3.1.4 Supply chain visibility

There have been several attempts to approach the concept of supply chain visibility in academia in order to define and explain it, for the purpose of this study the following definition of supply chain visibility was used: "a process of collecting supply chain data, integrating it, extracting intelligence from it, and using that intelligence to interrupt decisions in a cross-organizational supply chain oriented context." (McIntire, 2014, p. 25) (see figure 7).

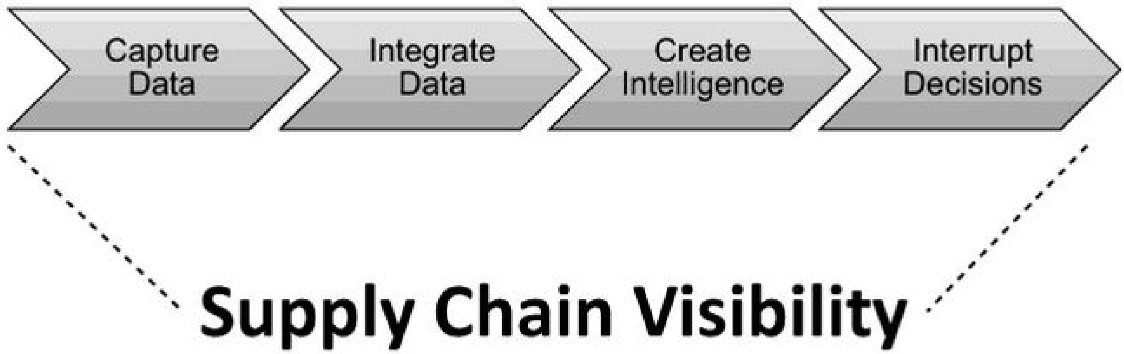


Figure 7- The four steps of Supply Chain Visibility (McIntire, 2014)

The first step of the process refers to the data collection element and especially in the way that data are collected and saved in the collecting systems used. Following the collection of data, there is a correlation or interconnection of the data collected in most of the cases for verification or identification purposes. This interconnection of the data collected leads to the third step which is the creation of intelligence out of it, meaning a meaningful output for the user derived from the interconnection of collected data. Finally the four step of the process takes place, in which the supply chain visibility solution will lead to an interruption of the decision process in most of the cases asking the user for actions to be taken based on the new intelligence created from the solution (McIntire, 2014).

3.2 Big data

In this section, the relevant definitions associated to big data are presented, as well as the business impact that different authors and organization have related to the use of big data.

3.2.1 Big data definition

According to Manyika et al. (2011, p. 1), “Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze”. Manyika et al. (2011) emphasizes as well in two aspects 1) the subjectivity of this definition, as it doesn’t describe big data in terms of a specific number of bytes or speed, but in terms of ‘typical tools’, converting big data in a moving target: what today is big data, tomorrow can be ‘normal data’ 2) The variability per sector, as the definition of big data will vary depending on the software tools most commonly used in a specific industry. A survey conducted by IBM (Schroeck, et al., 2012) on more than a thousand respondents reflects the broadness of the term big data (Figure 8). Nevertheless, authors in general agree on three dimensions that characterize big data: volume, velocity and variety (Zikopoulos, et al., 2012).

Volume

According to McAfee and Brynjolfsson (2012), in 2012 about 2.5 Exabyte of data, i.e. 2.5 billions of gigabytes, were created every day. In 2000, the total amount of data stored in the world was 800 Exabyte and it is estimated that by 2020 it will have grown to 35,000 Exabytes (Zikopoulos, et al., 2012). The reason for such big amount of data is mainly the technological advances that allow for companies to capture information in more detail than what they used to.

Velocity

McAfee and Brynjolfsson (2012), consider that the speed for collecting data is more important than the volume, as it allows companies to be more agile than competitors in their strategies. Velocity refers to the speed at which the data is flowing, as well as the short-life of data’s value (Zikopoulos, et al., 2012). In some cases, the information may arrive at high velocity, but it is possible to store it for later processing, while in other cases the data has to be processed quickly in order to make a fast decision on it (Dumbill, 2012). Examples of products used to manage this fast-moving data are IBM’s InfoSphere Streams, and the open source frameworks Twitter’s Storm and Yahoo’s S4 (Dumbill, 2012).

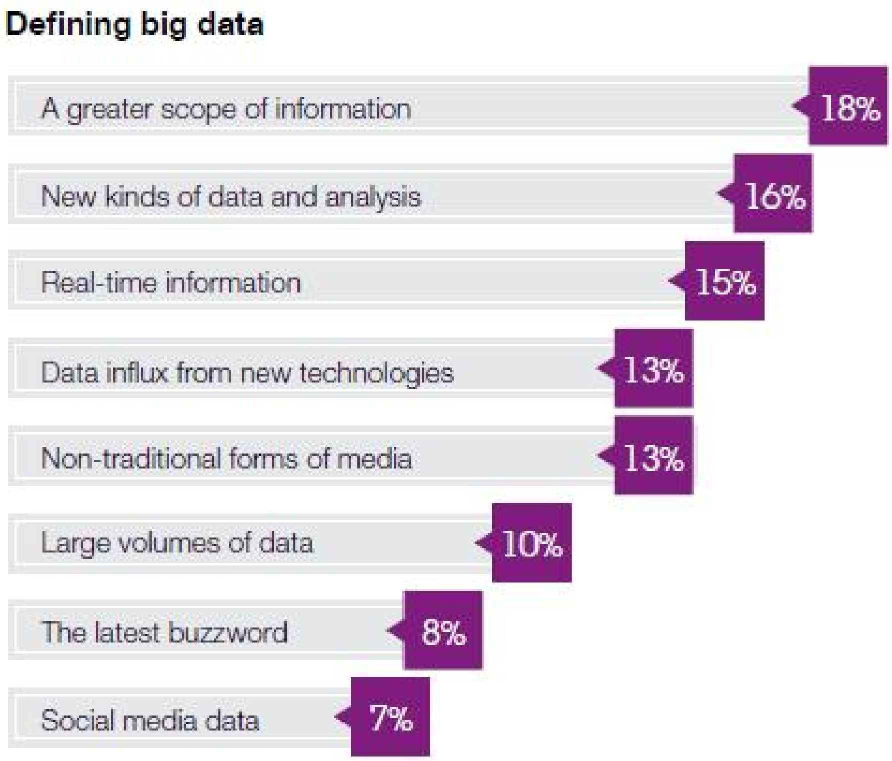


Figure 8 – How organizations see big data (Schroeck, et al., 2012)

Variety

In the past, information would come from one source or would be in a structured way. Nowadays, data is much more complex as it includes unstructured data from websites, social media, photos, web-log files, e-mails, documents, sensors, etc. (Zikopoulos, et al., 2012), and none of them are ready to be processed or integrated. According to Dumbill (2012), a big data application would structure the data and extract order meaning for future processing or analyzing.

3.2.2 Big data, big data analytics and data science

Another term that usually accompanies big data, is 'big data analytics'. Accenture defines it as: "the ability to gain insight from data by applying statistics, mathematics, econometrics, simulations, optimizations or other techniques to help a business make a decision about an issue or opportunity that needs to be addressed" (2014, p. 5). Based on this, it can be said that big data focuses on the technology for managing the data: extracting, storing, computing and visualizing it (Davenport & Patil, 2012), while big data analytics focuses on getting knowledge from it.

A third term that is very used nowadays is 'Data Science': even if it is not new (Press, 2013), it is in the last years that it has gained more importance due to its close relation to big data. According to Dhar (2013), Data Science focuses on extracting knowledge from the data i.e. understanding the key questions of the business, and getting the answers by employing mathematics, statistics, predictive modeling, computer programming and so on (Schmarzo, 2013). This definition is basically the same as for big data analytics, therefore, for the purpose of this study, the term big data analytics will be used to refer both big data analytics and data science.

Authors and the media in general, have a difficulties to differentiate the terms 'business intelligence' and 'big data analytics' or 'data science'. According to Schmarzo (2013, p. 38), they differentiate in the techniques and data types that they manage: business intelligence uses standard or ad hoc reporting, dashboards, alerts, queries, and basic statistical analysis to manage structured and manageable data. By the other hand, data science focuses on optimization, predictive modeling, forecasting, recommendations, and to achieve it, it uses advance statistical analysis, structured and unstructured data, as well as very large amounts of data.

3.2.3 Business Impact of big data and big data analytics

A lot of attention is being given to big data and big data analytics and how can they support business. The application and uses go from the medicine or criminology until supply chain and logistics. According to a survey done by Accenture, in the specific case of supply chain, companies have been able to achieve a significant number of results (see figure 9).

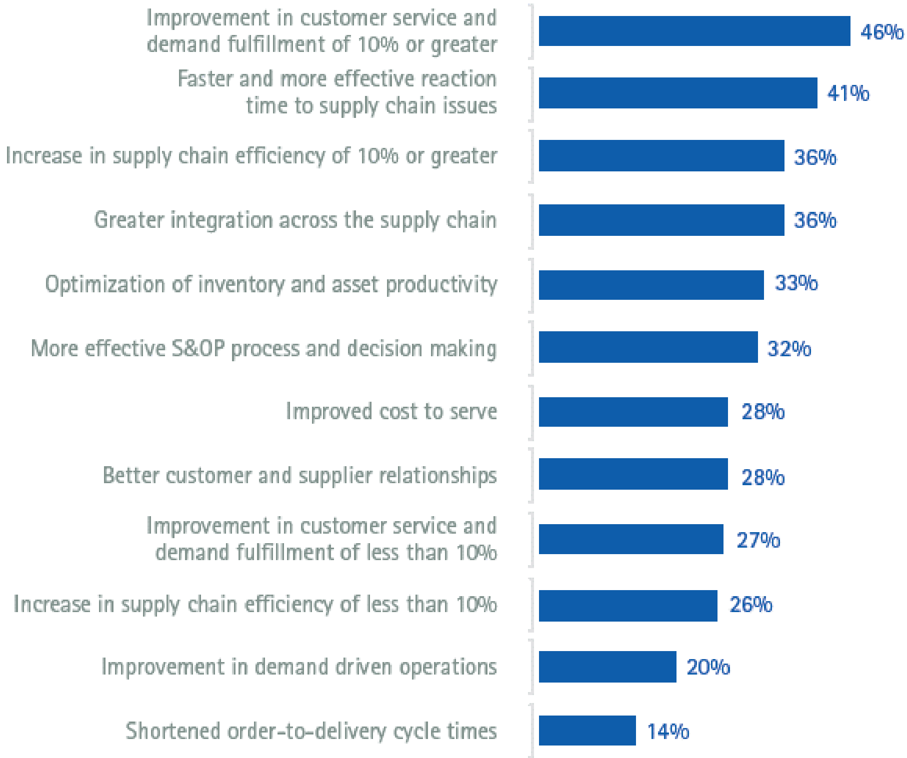


Figure 9 - Results companies have achieved using big data Analytics (Accenture, 2014)

After a revision of different authors, surveys and websites, the benefits of big data can be summarized as possibilities to predict the future, improve operational efficiency, better understanding and new business opportunities. Clearly, these benefits are inter-related and, in some cases, it is necessary to develop them in parallel to achieve the expected results.

Predicting the future

According to Sanders (2014). This is considered as one of the most valuable uses of big data analytics: the possibility to foresee or predict an event through the detection of small changes based on the analysis of current or historical data. One practical application would be the detection of the wearing of a part in a machine, car or a system, and determining its maintenance (Sanders, 2014). Another example of this, would be the 'anticipatory shipping' system recently patented by Amazon: based on an analysis of previous orders and other factors, Amazon is able to anticipate that a product will be bought, being able to ship it in advance and delivery it almost immediately after the customer has placed his order (Kopalle, 2014).

Operational efficiency:

The velocity factors of big data, play an important role when trying to reach operational efficiency (Jeske, et al., 2013). Through rapid processing of real-time information, DHL is currently being able to re-route their vehicles and re-define the delivery/picking sequence in order to save time (Jeske, et al., 2013); additionally, DHL has developed 'MyWays': a crowd-based platform that improves the efficiency of the last-mile delivery by assigning packages-to-be-delivered to commuters, students, taxi drivers based on their geo-location and usual routes (Jeske, et al., 2013).

Better understanding:

According to Sanders (2014), the use of big data and big data analytics can help to better understand what has happened and why things happen: the use of statistics, correlation analysis or predictive modeling is not new; what is new is the big amount of data that is possible to retrieve from internal or external sources, and the capability of the software tools that are able to process it. The result is a much better understanding of customer's behavior or preferences: for example, the capability of the store Target to identify when a customer is pregnant based on the change of their consumption patterns or how banks are able to detect fraud based on the abnormal use of a credit card (Sanders, 2014).

New business opportunities

Nowadays, the data collected is not only an asset for the owner of it, but for other companies as well (Sanders, 2014; Jeske, et al., 2013). Examples of this is how valuable geo-location data from mobiles has become, enabling marketing companies and departments to have a better understanding of how people moves within an area: at what time, how frequently and, based on that, decide on the location of stores or even send real-time advertisement based on users preferences and location (Gurbaksh, 2013). Another example of this, is the new product launched by DHL, Resilience 360 Incident Monitoring, a service provided to other companies where DHL provides real time information about incidents with the potential of affecting the supply chain of the customer; in this case, DHL is making use of their worldwide presence and the data gathered by itself or their partners (DHL, 2015).

3.2.4 Barriers for the success of big data

Sanders (2014) describes three main aspects of why the companies are not obtaining benefits from big data. These three aspects can act as barriers, or as enablers, depending on the degree of maturity or development.

Technology

Many companies have invested in advanced systems that create information silos within the organization, for example Customer Relationship Management (CRM) systems and manufacturing systems. In these cases, the data is isolated and the silos don't communicate with each other due to incompatible standards or formats. The upgrade of these systems is too costly, or sometimes impossible. To overcome this barrier, companies would have to invest on building bridges that connect those information silos or acquire new systems that can share information across the organization (Sanders, 2014).

People

Regarding people, there are mainly three aspects that can be considered as a barrier to implement big data: a) the lack of leadership: many companies still see data as a supporting input for decision making, without identifying its full potential as a competitive advantage; b) the lack of analytical talent: in many cases, the data is available but the people managing it don't have the analytical skills or statistics knowledge to convert the data into information and intelligence; c) culture: as the organization may be used to make decisions based on estimations or personal experience, instead of reliable data and facts (Sanders, 2014).

Processes

According to Sanders (2014), companies should focus not only on having the appropriate technology and people with the analytical skills, but on integrating both of them into the decision making process. Companies have what is called the 'siloed' decision making: actors within the supply chain make decisions without considering the effects up or downstream. Sanders (2014), suggests sales and operations (S&OP) and Collaborative Planning Forecasting and Replenishment, (CPFR) as the ideal process to implement big data analytics, as they are cross-functional and cross-enterprise.

4 Empirical Findings

4.1 Case framework

The logistics department of the focal company is responsible for the inbound and outbound transportation of material, products and spare parts. It is an organization with a footprint in 60 locations around the world employing around 5000 people.

The scope for the department includes making sure that material is transported to the production facilities, that packaging is available, that vehicles are distributed to the dealers, and that management of material, warehouses and distribution centers ensures the availability of parts everywhere in the world.

4.1.1 Material transport network in the focal company

The process that will be the primary focus during this research is the transport material process, i.e. the transport of materials from the suppliers to the factories. The focal company aims at ensuring that the materials are within three days distance from the factory in order to reduce the risk of material shortage. The main challenges of the process are to move the parts at the right time, creating the smallest stock possible, and at the lowest cost.

The focal company uses three means of transportation: road, which is mainly done by truck with contracted carriers; ocean container, used for long distance suppliers with low cost materials; and air, used for long distance suppliers with high cost components.

There are also two kinds of flows: direct flow, which refers to the movement of goods directly from one place to another, and cross docking, which refers to the sorting and reloading of the goods. Which kind of flow will be used, depends on the distance, the destination and the volumes. For example, in some cases, the suppliers are located near to the factory, and the material can be transported through a direct flow in a full truck load (FTL), a less than a truck load (LTL) or a milk run; in some other cases, the suppliers are too far or their volumes are too low, and a cross dock may be used.

4.1.2 Five Domains of the material transport process

The material transport process can be studied in five different domains as mentioned in the Introduction based on the decisions that are made and its impact in time: network design support, network configuration, planning and optimization, execution and analytic insight (see figure 10).

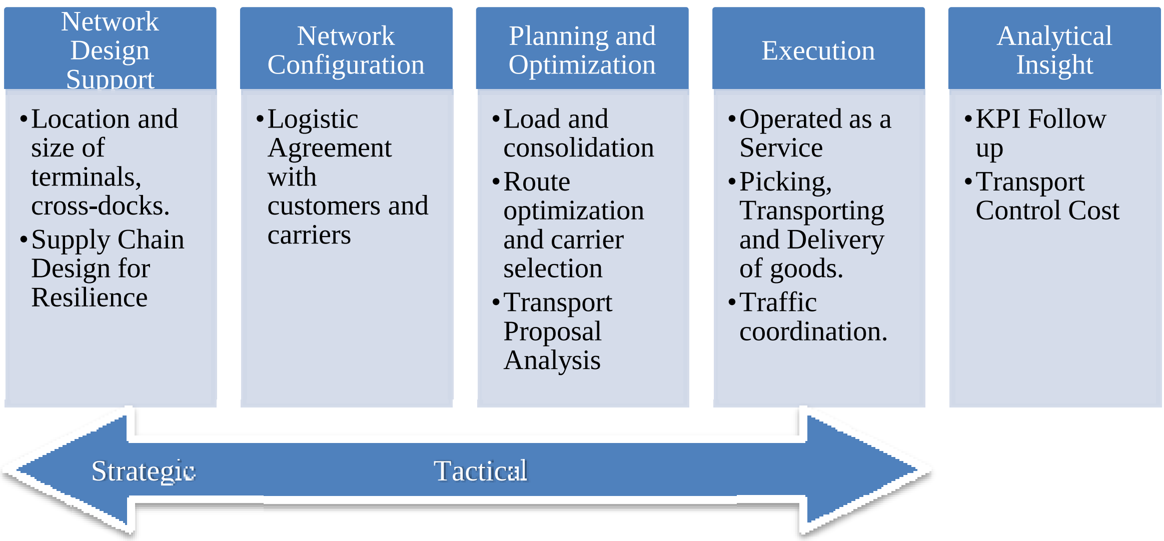


Figure 10 - Five domains in the transport material Process

Network design
The objective of this domain is to structure the foundations of the transportation network. The activities that fall under the network design when it comes to the transport material process are done in a strategic level which corresponds to a certain number of years. The logistics department is not responsible for the decisions regarding the location, number or size of the nodes of the network which in this case are the warehouses and the centers but is responsible to address the feasibility of these three given parameters from a logistics point of view. The responsible team that addresses this feasibility is called LCA (Logistics Cost Analysis).

Network configuration
The objective of this domain is to structure the connections between the different nodes of the transportation network previously defined by the network design domain as an output. The activities that fall under the network configuration are done in a tactical level in most cases in a yearly or bi-yearly frequency. There are two teams that cooperate under the network configuration domain: the purchasing team, responsible for handling the transport agreements every year or every two years, and the optimization team, responsible for the routing of the transportations and the optimization.

A cross-functional team decides upon the best possible routes to be followed between the plants, the warehouses or cross docking stations and the suppliers of material, based on the volumes and the prices. The purchasing team publishes the transportation needs in the form of RFQs (Requests for Quotation) to the transportation suppliers and request for offers.

Planning and optimization (Tactical)
The objective of this domain is to plan and optimize the possible routes that are given from the network configuration domain as an output. Optimization related activities are conducted in a tactical level which corresponds to yearly or quarterly frequency. The team that works with optimization in the logistics department is TNO team. Their primary objective is to optimize the network cost which in the case of transport it consists of cost elements like: costs associated with freight, storage and packaging.

They try to reduce these cost elements by setting the driving parameters like: the mode of transport to be used, the route to be followed, the frequency of pickups of the shipments, the lead time of the transport, the volumes of the materials and the pickup days. Some of these optimization decisions are executed via an in-house developed tool called MRT (Market Review Tool) which processes input data like past volumes (for example the volumes of the last eight weeks) and prices from the purchasing database and generates a plan, which is further optimized manually by the operator of the system based on their experience.

Planning and optimization (Operational)
The objective of the operational element of this domain is the balancing of the shipments and assignation to carriers. These activities are conducted in an operational level which corresponds to daily frequency. There is no dedicated team that executes this type of planning. An in-house built system executes this planning.

In order to explain this planning phase one needs to know the inputs of the planning. When there is a need for material, the material suppliers receive an order from the material planners working at the plants of the focal company. The order is sent in a message form. The next step is requesting the transportation of the material to the logistics department. This is done one day in advance, at least. The material suppliers make a TBR (Transport Booking Request) in the in-house built system.

During the planning process the in-house built system transforms the TBRs to TBs (Transport bookings) and assigns the shipment to carriers based on settings defined in planning and optimization tactical stage. The materials can either be shipped: a) directly from the material supplier's plant to the focal company's plant b) from the material supplier's plant through one cross docking station to the focal company's plant c) from the material supplier's plant through two cross docking station to the focal company's plant. These TBs are the outputs of the operational planning process and they are sent to the transport suppliers to be confirmed based on their capabilities. When the TBs are confirmed by the transport suppliers the execution of the transport of materials can begin. In case the TBs are rejected by the transport supplier then a retendering of the booking needs to be done until it is confirmed.

There are two more elements that affect the TB after it is issued by the system and before it can be executed. These elements are connected to packaging and premium transports. The packaging of the materials is handled by another system that aims to secure packaging for the materials to be transported. The premium transportations are mainly express shipments or shipments that require ad-hoc treatment.

Execution

When the TBs are issued to the transport suppliers, the execution phase of the transport of materials begins. All of the transports are executed according to the TBs and the transport planning. When the transport supplier picks up the shipment of materials from the material supplier's plant, the material supplier prints a POC (proof of Collection) document and both parties sign it. The POC must be registered to the in-house built system mentioned above no later than 2 hours after the collection window.

After that, the goods are transported according to the planning and the TBs directly to the receiving plants or through cross docking points. If any problems occur during transport TS (Transport Status) is issued in the in-house built system so that proper actions can be taken.

When the transport supplier reaches the gate of a receiving plant or a cross docking point a POA (proof of Arrival) message is issued by the receiver in the in-house built-system that is used to calculate transport material arrival precision on a shipment level. When the shipment is unloaded the receiver issues a POD (proof of Delivery) document in the in-house built system that is used to calculate the transport material arrival precision on part number level.

If the transport follows cross docking points then the same documents and messages are generated as described above for the unloading and loading procedures but there are some differences:

1. When the shipments reach the receiving area of the cross docking point, under special circumstances (overseas shipments, custom handling needed etc.) They are handled by two separate systems, not the in-house built system. This change of registration from the in-house built system to the two separate ones can be done automatically or manually by the cross docking operators.
2. After the first unloading at the cross docking point the TB that was used for the incoming shipment at the cross docking point is terminated and a new TB is generated for the outgoing shipment after the loading of the goods to leave the cross docking point.
3. If any problems occur during handling of material at the cross docking point a modification message needs to be issued so that the deviations are reported.
4. When the loading of goods according to the new TB is done, a LC (Load Confirmation) is issued in the in-house built system and matched with the planned departure time. In case of the special circumstances mentioned above this the updated TB will also be registered to the separate systems after the in-house built system is updated.

When the materials reach the final receiving point and the final POD is issued then the transport material process ends.

4.2 Material Forecast

The material planning process is performed in each of the assembly plants of the focal company: the sales orders and the sales forecast are processed in internal systems that, based on different parameters like safety stock, production date, bill of materials (BOM), packaging, and lead time, among others, determine the quantity and delivery date for each of the parts to be ordered. The result of this process is the material forecast.

It is the material controllers in the plant who are in charge of the material planning process: they make sure that suppliers receive the orders, follow-up that the delivery is on time and determine actions to be taken in case of delay, based on the current stock of the part.

The frozen window, or time when rescheduling of the order is not allowed, varies between suppliers and parts, and goes from two to thirty days.

4.3 Packaging Instructions

The focal company has developed its own packaging materials and packaging catalogue in a way that it suits the variety of materials (part numbers) that they require. Globally, the material suppliers use cardboard boxes, plastic boxes, wood pallets, wood frames and separators, among others, designed and supplied by the focal company. By doing so, the focal company can assure the availability of tools, machines, and space required to move or store the material in the plants.

It is the packaging developers in the plants who develop what is called 'The Packaging Instruction': a set of documents that describe how a specific supplier will pack a specific part number for a specific plant. In the packaging instruction, it is defined what type of internal boxes will be used to pack how many pieces, as well as what external packaging will be used, and what is the volume and gross weight of such THU. The volume and gross weight are calculated by the packaging developers based on the weight and volume of the packaging material and the parts to be shipped.

The packaging instructions are communicated to the material suppliers, who approve them or request modifications in the necessary case. All the packaging instructions are stored in a global system, where users can consult how a part number must be packed. It is important to mention that these packaging instructions and the volume and gross weight estimations are not an input information or data considered for any of the material transport process described in section 4.1.

5 Pre-Analysis

Based on the information gathered during the empirical study, it was possible to perform a pre-analysis and correlate the theoretical framework related to supply chain and logistics, with the domains in material transport process. The aim was to formulate a set of suggestions to the focal company in order to address the problems or gaps that were identified during the investigation. This set of suggestions, named as big data opportunities were formulated by analyzing the strategy of the focal company. In this chapter, the results of such pre-analysis are presented.

5.1 Identifying potential uses for big data

As described by Chopra and Meindl (2013), the two main drivers of the supply chain strategy are responsiveness and cost. The pre-analysis was performed through the use of cause-effect diagram, by identifying which causes would provide the desired effect: improved responsiveness and cost in terms of inventory management, transport, information and sourcing. The desired causes need to be related to the characteristics of big data: volume, velocity and variety. By doing so, fourteen big data opportunities were identified (Figures 11 and 12): some of the ideas support different drivers at the same time; nevertheless, for simplicity they are presented in the driver where they contribute the most.

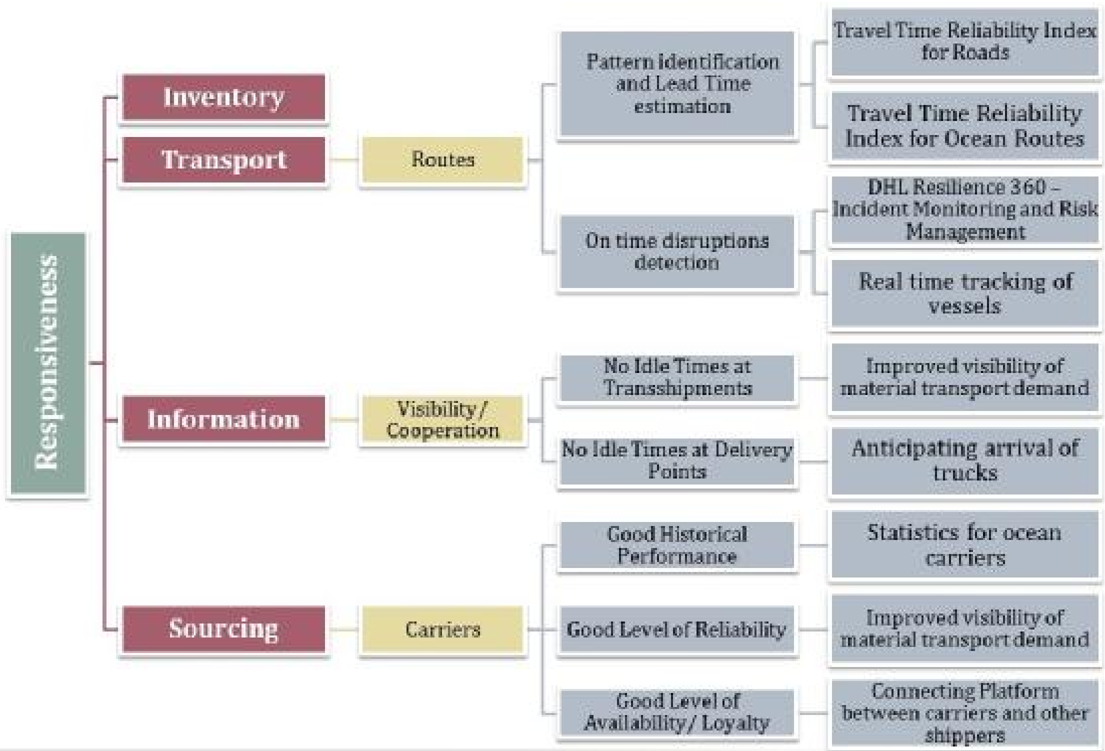


Figure 11 – Identifying big data opportunities to support responsiveness

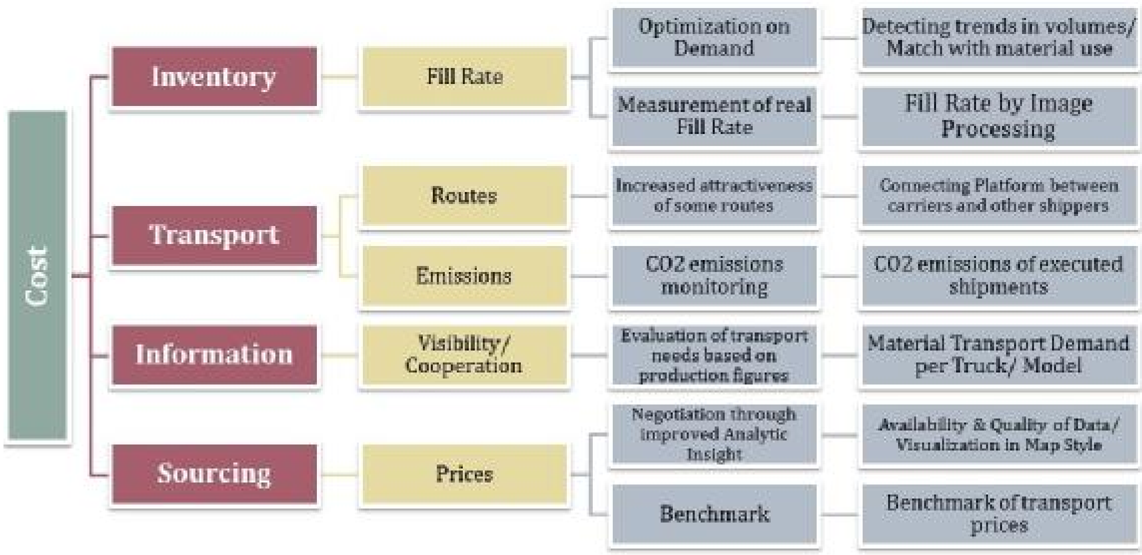


Figure 12 – Identifying big data opportunities to support responsiveness

By implementing any of the following big data opportunities the focal company can affect the responsiveness driver of the supply chain strategy:

1. **Travel time reliability index for road**: This opportunity concerns the creation of a reliability index in an effort to measure the overall systemic reliability of the road routes. This index would aid in the identification of variation patterns in the use of different routes as well as in the estimation of variations in the expected lead time.
2. **Travel time reliability index for ocean:** Similarly to the index described above this index would aim to measure the overall reliability of the ocean routes.
3. **DHL resilience 360:** This big data opportunity is an already existing solution in the market that can be applied in order to establish real time monitoring of incidents in transports as well as risk management suggestions. This solution can aid the effort to detect any disruptions in the transports on time.
4. **Real time trucking of vessels:** Another already existing solution in the market to positively affect the on time detection of disruptions. This solution can provide the focal company with weather information in the ocean as well as tracking of vessels.
5. **Statistics for ocean carriers:** Gathering of statistics from various data sources regarding the performance of each ocean carrier can be another application of big data aiming to establish a way to evaluate the historical performance of each carrier based on a large database of statistics.
6. **Improved visibility of material transport demand:** This big data opportunity aims to increase the overall visibility in the transport material process. Instead of booking transports based on TBRs from the material suppliers, the focal company can have the ability to directly translate the material orders issued by the plants towards the material suppliers, into weight, volume and quantity information in order to book the transports with the carriers. This effort can also affect positively the overall reliability of the transports executed by the carriers due to better input quality. In addition, by implementing this suggestion the focal company can also increase the reliability in its operations and more specific at the transshipment points at the terminals due to the increased visibility in the process.
7. **Connecting platform between carriers and other shippers:** This big data opportunity lies within the strong position of the focal company as a shipper. This strong position can be used for the creation of a platform to connect the serving carriers to the focal company with other possible shippers in an effort to increase the attractiveness of particular routes. By creating this platform the focal company is able to increase its security and also the reliability of the serving carriers because the probability of a TB to be rejected by a carrier decreases since the routes become more

attractive to the carriers. Moreover, the freight cost also decreases since the focal company can benefit from an increase in the competition among carriers due to the increase in the attractiveness of the routes.

8. **Anticipating arrival of trucks at delivery points:** By the use of GPS applications the arrival time of trucks can be estimated more accurately at the delivery points, thus increasing the reliability in the operations. By anticipating the trucks the waiting times can be significantly increased as well as the administration needs will be reduced since the associated papers like POA (proof of Arrival) will be replaced by electronic proof of arrival.

The focal company can affect cost driver of the supply chain strategy by implementing any of the following big data opportunities:

9. **Availability & quality of data/ visualization in map style:** This big data opportunity lies within the power of big data tools to visualize historical data like volumes and costs in a map style in order to aid the personnel responsible for the analysis. In addition, the overall insight is improved which allows better decision making. This in turn affects positively the negotiation process with the carriers since it will be based on improved analytic insight, thus decreasing the freight cost.
10. **Detecting trends in current volumes / match with the material use:** Freight cost can also decrease in terms of fill rate optimization that can be done on demand. Instead of executing the optimization of the volumes in a fixed period, optimization can be triggered on demand by continuous monitoring and trends detection in the volumes. Furthermore, if the volumes are matched with the actual material consumption then the lead time can be adapted accordingly resulting in a better control of the inventories, thus the costs associated with them.
11. **Fill rate by image processing:** Fill rate optimization can also be done by processing pictures taken with suitable applications in order to calculate more accurately the real filling degree.
12. **Benchmark of transport prices:** Freight cost savings can materialize by benchmarking of transport prices with the use of big data to access multiple data sources affecting capacity or rates. By doing that, real time follow up of the price levels in the carriers market can be done, thus aiding the purchasing process when tendering or negotiating with the carriers.
13. **Material transport demand per truck/ model:** Another big data opportunity aiming to increase the level of visibility in the transport material process. By combining production figures, BOM information and material orders per supplier, the focal company is able to produce the overall transportation needs per plant/truck. This will aid the process of evaluating the needs and the costs of transports as well as to analyze the possible impacts of changes in the production on the transports.
14. **CO2 emissions of executed shipments:** An application can be used to combine information regarding the model of the trucks and its emissions profiles with internal data from TBRs and TBs like distance, transport configuration and weight. The result will be a more accurate calculation of CO2 emissions of the executed shipments.

5.2 Big data opportunities in the five domains

The big data opportunities previously identified were analyzed considering the five domains of the material transport process: which ideas would support the focal company in a strategic, tactical and operational level. As shown in figure 13, it was identified that some of the opportunities will support more than one domain.

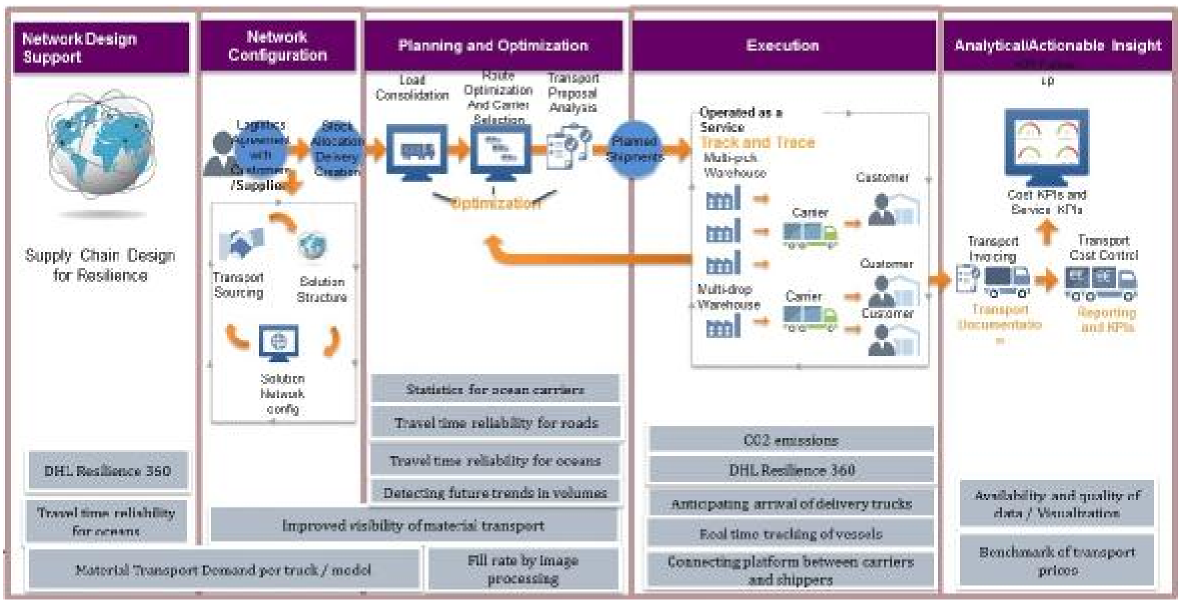


Figure 13 - Big data opportunities in the five domains of material transport process

The benefits of the big data opportunities identified have been considered in order to be reviewed in the next section, the development of the proof of concept.

6 Proof of concept

In the previous chapter, pre-analysis, fourteen big data opportunities that would support the material transport process were identified. With the aim evaluating the difficulty, barriers and benefits of using big data, one of the ideas was developed into a proof of concept. This chapter presents the proof of concept developed: from the selection process until the assessment of the model. To ease the understanding of the development and findings, the information is presented following the phases of the selected methodology, CRISP-DM (figure 14).

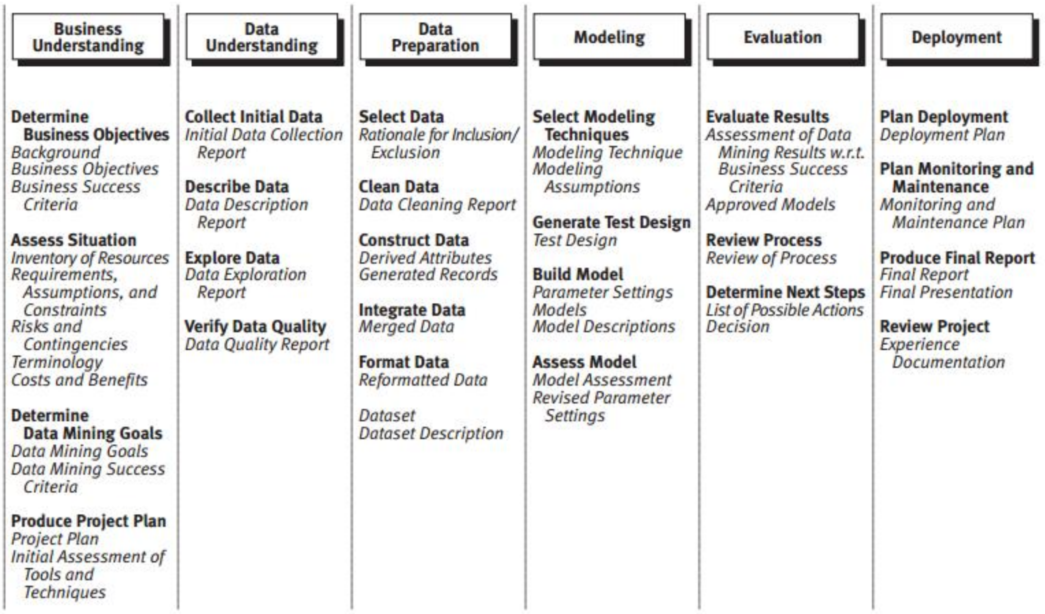


Figure 14 - Phases and general tasks of the CRISP-DM Methodology

6.1 Business Understanding

According to Chapman et al. (1999, p. 10), the initial phase of the CRISP-DM Methodology "focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition". With this aim, some generic tasks are performed: Determining business objectives, assessing the situation through an inventory of resources and a cost benefit analysis, determining the goals and producing a project plan.

6.1.1 Determine Business Objectives: Prioritizing big data Opportunities

In order to align the big data Opportunities detected with the business objectives and strategy, an anonymous web survey was designed and sent (See Appendix 1). The respondents were selected based in a way that they fulfilled at least one of two requirements: a) they are key decision makers within the material transport process at a strategic level, b) the tasks they perform are directly related to one of the five domains of the material transport process. The respondents were requested to evaluate each of the opportunities and its possible benefits in a scale from 1 to 5, following the classification guidelines shown in Table 1 – classification of benefits.

Table 1 – Classification of benefits

BENEFITS			
1	No benefit		The solution has no impact on KPIs: - Doesn't address any current problem, or - doesn't improve delivery times, or - doesn't provide insight for better decision making, or - doesn't save money or time.
2	Low Benefit		The solution has low impact on KPIs: - Addresses a sporadic problem with low financial consequences, or has small effect on delivery times, or - provides little improvement on insight for decision making, or saves small amount of money or time.
3	Medium Benefit		The solution has medium impact on KPIs: - Addresses a recurrent problem with medium financial consequences, or - has medium effect on delivery times, or - saves considerable amount of money or time, or - provides considerable improvement on insight for decision making.
4	High Benefit		The solution has high impact on KPIs: - Addresses a frequent problem with high financial consequences, or - has a major effect on delivery times, or - reduces an important amount of money or time, or - improves the decision making by providing a new perspective.
5	Very High Benefit		The solution has a very high impact on KPIs: - Addresses a critical problem with high financial consequences, or - will highly improve the delivery times, or - reduces a big amount of money or time, or - improves the decision making by providing a new perspective or time.

From the 23 surveys sent, 13 answers were received, i.e. a response rate of 56. With the aim of eliminating the possibilities of a result biased by the number of respondents per department, an ‘average benefit index’ was calculated considering the average per department. The benefit scale identified from the survey is as shown in table 2.

Table 2 – Results for benefits survey

		Controlling	LCA	OD/VPS	Operations	Optimization	Purchasing	Average Benefit
	Number of surveys sent	1	4	1	5	5	7	
	Number of respondents	1	3	1	2	2	4	
1	Anticipating ArrivalTrucks	5,00	4,00	3,00	4,00	3,50	3,75	**3,88**
2	CO2 Emmissions Executed Shipments	5,00	4,00	2,00	3,00	2,50	2,75	**3,21**
3	DHL Resilience	4,00	4,00	3,00	3,50	2,50	3,25	**3,38**
4	Travel Time Reliability Index for Ocean	4,00	4,33	2,00	3,50	3,00	3,50	**3,39**
5	Statistics for Ocean Carriers	4,00	3,67	4,00	4,00	3,00	3,50	**3,69**
6	Real Time TrackingVessels	4,00	3,67	3,00	4,00	3,00	3,25	**3,49**
7	Travel Time Reliability Index for Road	4,00	4,33	3,00	4,00	3,50	3,50	**3,72**
8	Improved Visibility Material Transport Demand	4,00	5,00	N/E	4,00	4,00	4,50	**4,30**
9	Material Transport Demand per Truck / Model	5,00	4,67	3,00	4,00	4,00	4,25	**4,15**
10	Detecting Trends in current volume	4,00	4,00	3,00	3,50	3,50	2,50	**3,42**
11	Availability & Quality Data / Visualization	5,00	4,33	4,00	3,50	3,00	4,50	**4,06**
12	Platform Connecting other shippers with focal company's Carriers	4,00	3,33	N/E	3,50	3,00	3,50	**3,47**
13	Benchmark for Transport Prices	5,00	4,00	N/E	3,50	3,00	4,00	**3,90**
14	Fill Rate by Image Processing	4,00	3,67	1,00	3,50	3,00	3,50	**3,11**

6.1.2 Assess the Situation: Complexity and big data Index

The second task of the business understanding is assessing the situation. This was done in three steps: an inventory of the input data required for each of the big data opportunities; an assessment on the complexity through the evaluation of actions required; and the calculation of a big data index, developed by the researches, in order to identify the relevance of the opportunities as objects of a proof of concept.

The inventory of the input data required (Table 3) was done through: an analysis of the information from interviews, i.e. what users and systems provide relevant information for the opportunity, a research on the data services available in the market, and an analysis of different solutions and technologies used in the field.

Table 3 - Input data required for big data opportunities.

#	Opportunity description	Input data
1	Anticipating arrival of trucks with GPS/APP	- GPS Information from trucks: calculation of ETA. - Current information on the status of the terminal: availability of bays - Info from TB: TB Order, Carrier, Volume, Weight
2	CO2 Emissions of executed Shipments	- Characteristics from truck used for transport: type of vehicle and fuel used received via a mobile application. - Information with internal data for distance, load, etc. - Information about the routes (type of road: urban, motorway, etc.)
3	DHL Resilience 360	- Incidents tracked worldwide. - Vulnerabilities tracked worldwide: environmental, social, economic, political.
4	Travel Time Reliability Index for Oceans	- Historical data from ports: calls, turnaround times; compare Estimated turnaround vs actual turnaround (data provided by internet services).
5	Statistics for Ocean Carriers	- Historical data from carriers: ETA vs Actual Time of arrival, among others (data provided by internet services).
6	Real Time Tracking of Vessels	- Vessel monitoring (location + estimation of delay): data provided by internet service. - Info from internal systems: in which vessels material is being transported.
7	Travel Time Reliability Index for Roads	- Build historical data from Google Maps for certain routes. - Info from internal systems: which roads and routes are used.
8	Improved visibility of Material Transport Demand	- Material forecast from ordering system. - Volume / weight from packaging database.
9	Material Transport Demand per Plant / Truck Model	- Production forecast (plants). - BOM. - Volume / weight from packaging database.
10	Detecting Trends in Current Volumes	- Daily extraction from the transports systems.
11	Availability & Quality of Purchasing Historical Data / Visualization	- Daily extraction from the transports systems.
12	Platform connecting other shippers with focal company's carriers	- Available capacity from carriers. - Carrier performance.
13	Benchmark for Transport Prices	- Prices from different container vessels provider. - Follow-up of variables affecting prices.
14	Fill Rate by Image Processing	- Picture from trucks loaded. - Matching information with shipment data from the systems.

With the aim of evaluating the complexity of the big data Opportunities, the steps necessary to achieve the implementation of the opportunities were identified. Six main tasks were identified in this respect: a) Data/Service Acquisition, which refers to the need of buying the service or the data from an external source; b) Data Collection / Cleaning, refers to the step of extracting the data from an internal or external source and cleaning it, i.e. removing the data that is not relevant; c) Matching/Comparison describes the need of merging two or more data sources in order to achieve the expected result; d) Analytical or Processing required, refers to the calculations or mathematical/statistical analysis that may be required; e) Organizational impact describes the effect that the opportunity would have on the users tasks and responsibilities, once it is implemented; and f) Software Development Required, refers to the need of developing an interface or algorithm in order to reach the solution. The evaluation of these points is presented in table 4. As it can be observed, two complexity indexes have been calculated: the first one, General Complexity, evaluates the six above-mentioned tasks in equal way, adding one point per task to calculate the index. The second one, short-term complexity, adds one point per task as well, nevertheless, due to time limitations and the nature of this research, an index ‘5’ was assigned by default in those cases where software development was required.

Table 4 – Complexity evaluation

#	Opportunity Description	Data/Service Acquisition	Data Collection / Cleaning	Matching/Comparison of two or more data sources	Analytical or Info Processing required	Organizational Impact	Software Development Required	General Complexity[1]	Short-term complexity[2]
1	Anticipating arrival of trucks with GPS/APP			1	1	1	1	**4**	**5**
2	CO2 Emissions of Executed Shipments		1	1	1	1	1	**5**	**5**
3	DHL Resilience 360	1						**1**	**1**
4	Travel Time Reliability Index for Oceans	1	1	1	1			**4**	**4**
5	Statistics for Ocean Carriers	1	1	1				**3**	**3**
6	Real Time Tracking of Vessels	1		1		1	1	**4**	**5**
7	Travel Time Reliability Index for Roads		1		1		1	**3**	**5**
8	Improved visibility of Material Transport Demand		1	1	1	1		**4**	**4**
9	Material Transport Demand per Plant / Truck Model		1	1	1			**3**	**3**
10	Detecting Trends in Current Volumes		1	1	1	1		**4**	**4**
11	Availability & Quality of Purchasing Historical Data / Visualization		1	1	1			**3**	**3**
12	Platform connecting other shippers with focal company’s carriers		1	1	1	1	1	**5**	**5**
13	Benchmark for Transport Prices	1						**1**	**1**
14	Fill Rate by Image Processing		1	1	1		1	**4**	**5**

The final step of assessing situation task was determining a big data index, i.e. how ‘big data’ is each of the fourteen opportunity. With this aim, the characteristics of big data, volume, velocity and variety, were evaluated for each of the opportunities’ input data. The scale, 1 to 3, was defined by the researches, according to table 5, and it ranks how big the volume of the data would be, how fast the

[1] *The complexity for the focal company was evaluated considering the number of activities that would take place in order to develop the project.*
[2] *Due to the nature of this study, those opportunities that required a software development were ranked as 5 due to the time and programming that would be required.*

data would lose its value, and the type and quantity of data sources required. The bigger the index is, the more 'big data' the opportunity is. The results of this evaluation are presented in table 6.

Table 5 – Guidelines for evaluation of volume, velocity and variety

	VOLUME	VELOCITY	VARIETY
1	Data can be managed by current software available within the focal company	Information loses its value within a year; decisions over the information can be made in the long-term.	One or two data sources with the same kind of format.
2	Data must be managed by more advance software (Qlikview, for example)	Information loses its value within a month; decisions over the information need to be made in a one-month period.	More than two sources of data are needed; but still the majority of the data comes from databases with tabular or columnar arrangements.
3	Data must be managed by specific big data Software and tools	Information loses its value immediately; decisions over the information need to be made immediately	The data has different forms: text, images, sensor signals, GPS positions, data from mobile applications, etc.

Table 6 - Big data index: evaluation of volume, velocity and variety for the big data opportunities

#	Idea Description	Volume	Velocity	Variety	Big data Index
1	Anticipating arrival of trucks with GPS/APP	1	3	3	2,33
2	CO2 Emissions of Executed Shipments	1	1	3	1,67
3	DHL Resilience 360	2	3	3	2,67
4	Travel Time Reliability Index for Oceans	2	1	1	1,33
5	Statistics for Ocean Carriers	2	1	1	1,33
6	Real Time Tracking of Vessels	1	3	2	2,00
7	Travel Time Reliability Index for Roads	2	2	2	2,00
8	Improved visibility of Material Transport Demand	2	2	2	2,00
9	Material Transport Demand per Plant / Truck Model	2	2	2	2,00
10	Detecting Trends in Current Volumes	1	2	2	1,67
11	Availability & Quality of Purchasing Historical Data / Visualization	2	1	2	1,67
12	Platform connecting other shippers with focal company's carriers	2	3	2	2,33
13	Benchmark for Transport Prices	2	2	2	2,00
14	Fill Rate by Image Processing	2	2	3	2,33

6.1.3 Determining goals

With the aim of determining the big data opportunity that will be carried to the further steps of the CRISP-DM Methodology, i.e. the development of a proof of concept, a summary of the previous tasks was done: a table and a chart summarizing the benefits, complexity and big data index of the opportunities is presented in table 7 and figure 15. The purpose was to identify one big data opportunity that had a good compromise between the three parameters: the highest possible benefit (vertical axis) for an acceptable level of complexity (horizontal axis), with the biggest data index possible (bubble size).

Table 7 - Summary of Benefits, Complexity and big data Index

#	Opportunity	Benefit	Complexity	Big data
1	Anticipating Arrival of Trucks	3,88	5	2,33
2	CO2 Emissions Executed Shipments	3,21	5	1,67
3	DHL Resilience	3,38	1	2,67
4	Travel Time Reliability Index for Ocean	3,39	4	1,33
5	Statistics for Ocean Carriers	3,69	3	1,33
6	Real Time Tracking of Vessels	3,49	5	2,00
7	Travel Time Reliability Index for Road	3,72	5	2,00
8	Improved Visibility of Material Transport Demand	4,30	3	2,00
9	Material Transport Demand per Truck / Model	4,15	3	2,00
10	Detecting Trends in current volume	3,42	3	1,67
11	Availability & Quality Data / Visualization	4,06	3	1,67
12	Platform Connecting other shippers with focal company's Carriers	3,47	5	2,33
13	Benchmark for Transport Prices	3,90	1	2,00
14	Fill Rate by Image Processing	3,11	5	2,33

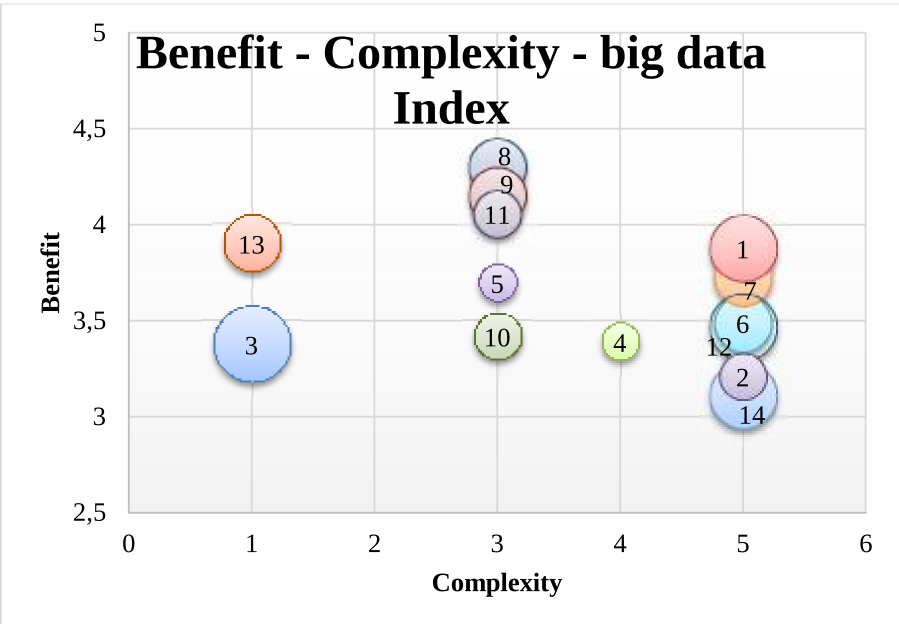


Figure 15 - Evaluation of big data opportunities: benefits (vertical), complexity (horizontal) and big data index (bubble size).

Based on the evaluations performed, it was identified that the big data Opportunity 'Improved Visibility of Material Transport Demand' would provide the highest benefits for the organization, while having an intermediate level of complexity and big data Index, being appropriate for the scope and aim of this study. A summary of the idea and the theoretical benefits are presented in figure 16.

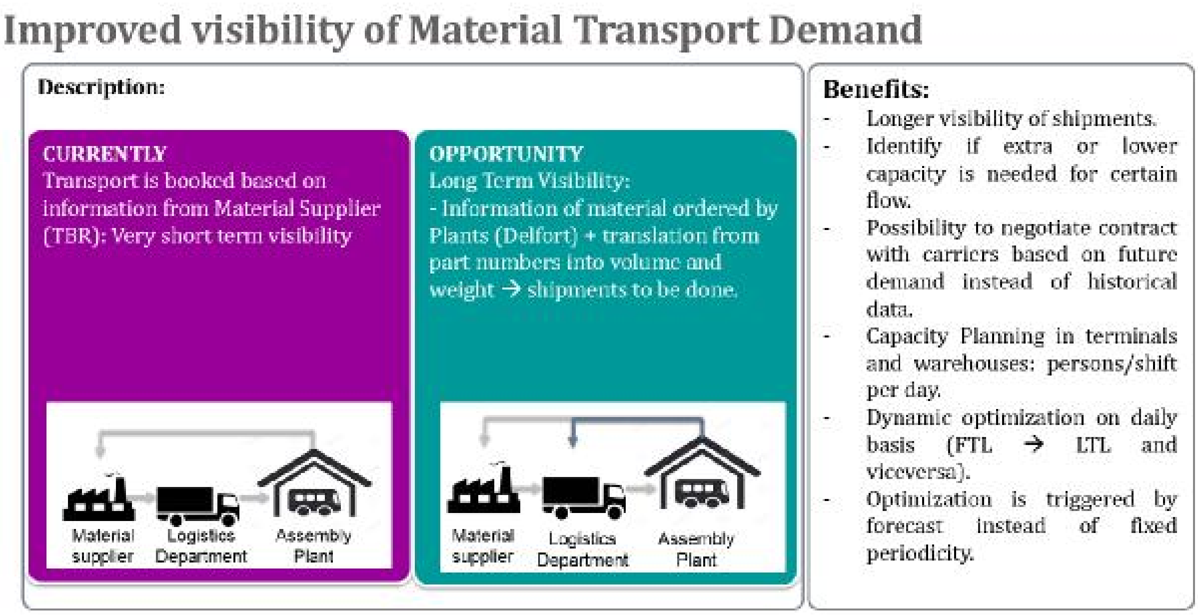


Figure 16 - Improved visibility of material transport demand

6.1.4 Produce project plan

In order to improve the visibility of material transport demand, it was decided to create, as a proof of concept, a material transport forecast for one plant, taking as input data: a) the material forecast from ordering system, i.e. the information created by material controllers in the plants, containing the quantity and delivery date of each part from each supplier; b) the packaging instructions for each part for each supplier, which would be used to calculate the volume and weight that each of those parts would require. The expected result of the proof of concept was to have the total of THU (Transport Handling Unit), weight (tonnes) and volume (m^3) to be delivered per country of origin, per week.

6.2 Data understanding

The second phase of the CRISP-DM process refers to the understanding of the data: collecting, describing, exploring and verifying it. As it has been observed previously, the CRISP-DM Methodology is not only one-directional, but allows to make adjustments in the previous phase, business understanding, after an initial understanding of the data.

6.2.1 Collection and description of initial data

Initially, it had been decided to make a proof of concept for one of the plants in Europe. Nevertheless, after an initial collection of data, it was identified that a significant quantity of information regarding the packaging instructions was missing. It was then decided to change the object of the proof of concept, and a different plant within Europe was selected.

The data that was collected for this purpose was:

a) Forecast of material Deliveries: excel file with information about material supplier number (sometimes called Original Consignor), part number, quantity to be delivered, week of delivery, customer (sometimes called Final Consignee).
b) Packaging instructions: data collected from a worldwide repository containing information about supplier number, part number, customer, units per load (the quantity of pieces that can be loaded into a THU), volume per THU, Gross weight per THU, as well as the type of external packaging, internal packaging and units per internal packaging (quantity of pieces that can be loaded into an internal box or bag).

c) Flows within Europe: excel file containing the location of the material supplier (original consignor), the location of the different plants where the supplier can ship from (first consignor), the delivery points (final consignee), the mode of transportation (full truck load or less than a truck load) and the pick-up days.

6.2.2 Data exploration and verification of quality

Each of the data sets was explored and verified separately, with the aim of confirming that all the needed information had been collected.

Forecast of material deliveries:

An exploration of the excel file as a source of data revealed that some of the required information was available: the quantity of parts to be shipped, the delivery week, and the delivery place. Nevertheless, it was identified that the file contained only the original consignor, i.e. the supplier name and number to which the order was placed, not the first consignor, i.e. the location or supplier's plant from where the actual shipment would be performed.

Packaging instructions:

The data was collected from the worldwide repository: an excel file was created containing all the packaging instructions from all the suppliers and all their part numbers delivered to the plant. After a first exploration of the data, it was identified that the majority of the packaging instructions contained the information required: volume (in cubic meters), the gross weight of the THU, and the quantity of pieces that fit into each THU. Nevertheless, three problems were identified: the first problem was that a significant number of packaging instructions didn't have all the required data: THU, volume and weight, i.e. the information was incomplete; after investigating on this issue with the packaging designers, it was concluded that the data was missing due to packaging flexibility, i.e. the THU can contain more or less pieces, depending on the order placed by the material controller. The second problem identified, was the unavailability of packaging instructions for specific part numbers: they couldn't be found in the system. The third problem was wrong data in the system: for example, some pallets or loads had been assigned with a too high gross weight, in the level of tens or hundreds of tonnes, which after an initial verification, resulted to be caused by a human mistake when entering the information in the system. The three problems were managed in subsequent steps of the methodology.

Flows within Europe:

The data of different flows within Europe was collected from optimization department whose responsibility is to optimize the balance within Europe. This data consists on the location data like address, city, postal code, etc., for the supplier (original consignor), the suppliers' shipping facilities (first consignor) and the customer plants (final consignee). The data was verified through a sampling against shipment system.

6.3 Data preparation

The purpose of the data preparation phase is to deliver a dataset that will be used in the modeling stage. It consists of several general tasks, mainly aimed at selecting, cleaning, constructing, integrating and reformatting the data in a way that can be managed in subsequent stage.

6.3.1 Select and clean data

After the understanding of the data in the previous stage, it was identified that there was no need to exclude, or remove, information from the files 'forecast of material deliveries' and 'flows within Europe', as the quality and amount of data was appropriate for this proof of concept. Nevertheless, in agreement with the focal company, it was decided to exclude the missing data related to packaging instructions (described in section 6.2.2), i.e. disregard those part numbers that had no existing packaging instruction, as there wasn't information to calculate the volume or weight to be transported. This was done mainly due to two reasons: 1) the data was not available in the system; 2) the amount of missing data was not relevant for the purpose of this proof of concept. It is important to note that the percentage of missing data cannot be revealed due to confidentiality agreements with the focal company.

Additionally, as it had been observed in section 6.2.2 there were packaging instructions with wrong information: for example, weight of the pallet of several tens or hundreds of tonnes); this data was manually removed from the system through a search for all the part numbers whose THU had a gross weight higher than 3500 kg., which was considered as above the average. Nevertheless, there is the possibility that other incorrect data, volume or weight, has been incorrectly calculated by the packaging department. Due to the amount of data, it was not possible to identify and remove other part numbers with incorrect data, which will lead to an unknown margin of error in the calculations of volume and weight.

6.3.2 Construct data

Regarding the part numbers whose packaging instructions had incomplete information (first problem of packaging instructions described in section 6.2.2) it was decided to construct the missing data: an estimation of the volume, weight and THUs to be transported is possible through calculations using additional information: a) the packaging catalogue, a PDF file containing the volumes and weight of the packaging elements like external pallets, wood frames, plastic boxes and cardboard boxes; b) the part weight, and c) the number of pieces that fit into the internal packaging. See Appendix 2 for the formulas of the volume and weight used to construct the missing data.

6.3.3 Integrate and format data

No formatting or integration of the data was necessary in this stage of the process.

6.4 Modeling

In this phase of the CRISP-DM Methodology a technique or tool is selected and applied, the settings are put into the system and, in the necessary case, an interface is developed.

6.4.1 Selection of modelling technique

The tool selected to develop the proof of concept was Qlikview as it was identified that it would have the capabilities to manage the type of input data that would be used, providing the results that this proof of concept was trying to achieve. Qlikview is “a platform for creating and deploying rich, guided analytics applications via a native scripting environment (…) that enables users to intuitively explore data relationships across many sources that would be hidden in hierarchical or query-based approaches” (Qlik, 2015).

6.4.2 Generate Test Design

In order to verify the quality and validity of the forecast calculated, the following test were defined (table 8):

Table 8 – Tests for proof of concept

	Description	Purpose	Sampling	Expected result
Test 1	Compare information from executed shipments (in transport booking system) with proof of concept	Verify that the calculation or estimation of THU, weight and volume, are correct	24 shipments[3]	The calculations are within a range of +/- 10%[4].
Test 2	Take the forecast of materials created in week 10, 2015 and process it in the proof of concept. Compare with THUs, volume and weight actually shipped in weeks 11 to 18 (information available in the system for historical data).	Verify that the proof of concept is capable of detecting peaks and trends.	Weeks 11 to 18	Peaks and trends in both calculations are similar.

6.4.3 Build Model

The model was built based on the data from the three different files previously described. It was set in a way that the keys of the system, i.e. the unique parameters that link one data source with another, would be the supplier number, the part number and the delivery plants. The system, Qlikview, created synthetic keys by identifying by itself such unique and common data within the data sources. A diagram of the links and information is shown in Figure 17. For detailed information on how the information was loaded, see the script in Appendix 3.

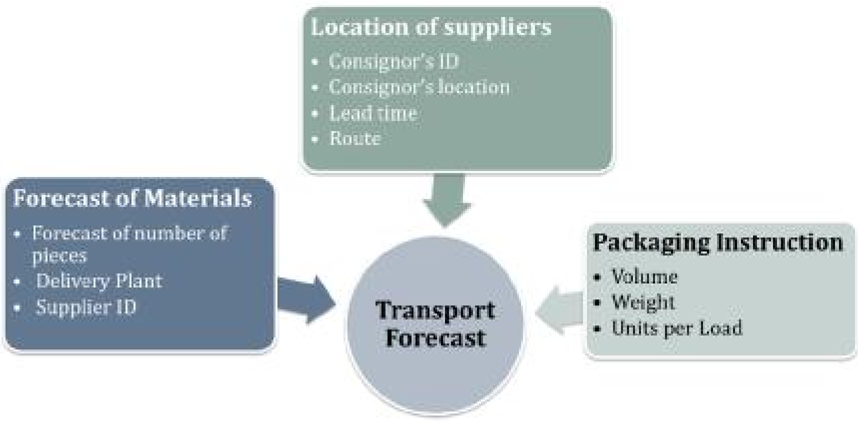


Figure 17 - Diagram of data in Qlikview.

Based on the data that had been collected, a model was built in the form of a dashboard with information that was considered as key for the possible users (see Figure 18): total THUs, Weight, Volume, as well as a chart representing how these will behave during the next twelve months and the geographical source. The dashboard was built as an interactive interface, where the users can gain a deeper understanding of the numbers by clicking in the areas where they would like to get more

[3] Based on ANSI/ASQ Z1.4, "Sampling Procedures and Tables for Inspection by Attributes", General inspection level II, AQL 1.0.

[4] No previous similar studies have been done before in this regard within the automotive company (theoretical weight vs. real weight), therefore there are no upper or lower control limits as reference. As consequence, an initial limit of +/-10% was set in order to have a parameter and categorization.

information, for example, supplier number, part number, country of origin, week number or picking days.

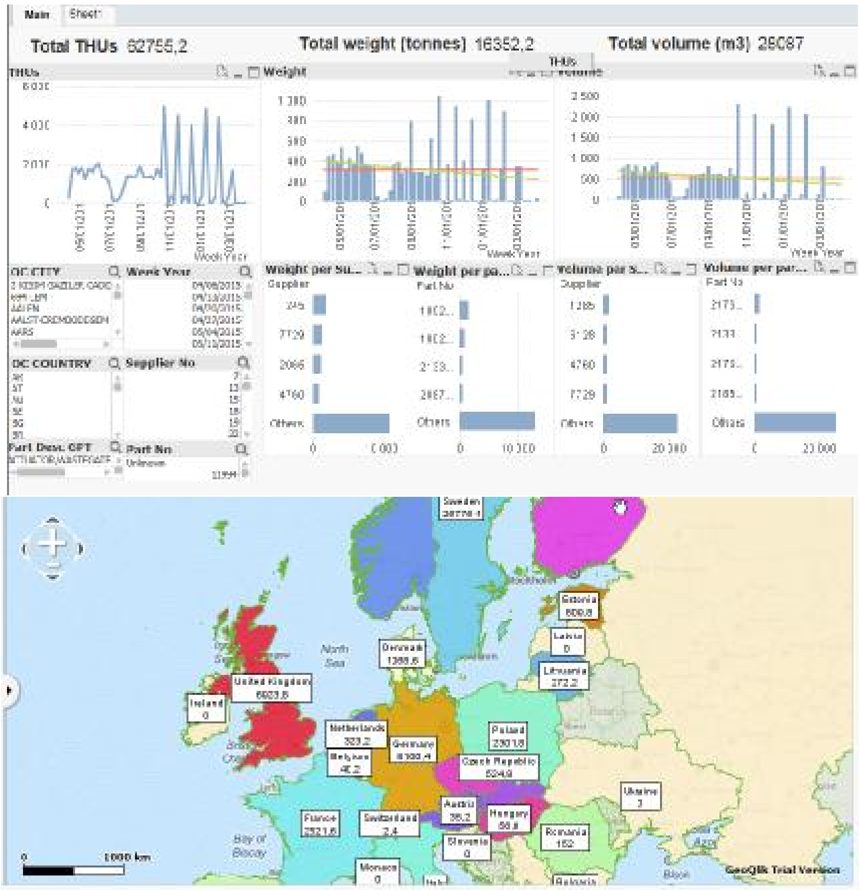


Figure 18 - Proof of concept: material transport demand for one assembly plant

6.4.4 Assess Model

The final task of the modeling phase, refers to assessing the technical success of the application, summarizing the results and listing the quality of the generated model.

Test 1

The results obtained in test 1 were not the results expected: it was identified a mismatch between the expected THUs, volume and weight, and the actual THUs, volume and weight for different shipments performed. A comparison between them is presented in table 9.

After reviewing the results of the 24 shipments, it was possible to identify that there were differences in terms of quantity of THUs, volume and weight. For THUs, in 79% of the cases the quantity shipped matched with the forecast. For the weight and the volume, in some of the cases the weight per THU shipped was within the tolerance of weight expected +/- 10%; while in other cases, the weight per THU shipped was higher or lower than the forecast by more than 10%.

Table 9 - Results of test 1

Proof of Concept - Translation of material forecast						In-house built system - Real shipments					
Week	THUs	Weight	Volume	Weight/ THU	Volume/ THU	Week	THUs	Weight	Volume	Weight/ THU	Volume/ THU
W16	3	0,356	0,33	0,12	0,11	W16	3	0,175	0,342	0,06	0,11
W17	11	0,388	0,594	0,04	0,05	W17	3	0,110	0,13	0,04	0,04
W17	2	0,072	0,108	0,04	0,05	W17	1	0,036	0,054	0,04	0,05
W16	1	0,010	0,034	0,01	0,03	W16	1	0,022	0,038	0,02	0,04
W17	23	0,364	1,242	0,02	0,05	W17	25	0,550	1,236	0,02	0,05
W16	1	0,006	0,034	0,01	0,03	W16	1	0,001	0,004	0,00	0,00
W16	5	0,033	0,27	0,01	0,05	W16	5	0,006	0,028	0,00	0,01
W16	3	0,051	0,036	0,02	0,01	W16	3	0,077	0,228	0,03	0,08
W17	1	0,128	0,148	0,13	0,15	W16	1	0,126	0,154	0,13	0,15
W17	1	0,129	0,148	0,13	0,15	W16	1	0,126	0,154	0,13	0,15
W16	2	0,009	0,068	0,00	0,03	W16	1	0,010	0,034	0,01	0,03
W17	3	0,068	0,102	0,02	0,03	W17	2	0,046	0,108	0,02	0,05
W16	4	0,102	0,896	0,03	0,22	W16	4	0,064	0,936	0,02	0,23
W16	3	0,096	0,444	0,03	0,15	W16	3	0,090	0,462	0,03	0,15
W16	1	0,010	0,034	0,01	0,03	W16	1	0,024	0,034	0,02	0,03
W16	2	0,172	0,088	0,09	0,04	W16	2	0,180	0,936	0,09	0,47
W16	1	0,025	0,034	0,03	0,03	W17	1	0,025	0,034	0,02	0,03
W16	3	0,049	0,102	0,02	0,03	W17	3	0,050	0,102	0,02	0,03
W16	1	0,005	0,034	0,00	0,03	W16	1	0,004	0,036	0,00	0,04
W16	1	0,013	0,054	0,01	0,05	W16	1	0,017	0,074	0,02	0,07
W16	1	0,009	0,034	0,01	0,03	W16	1	0,009	0,034	0,01	0,03
W16	1	0,022	0,054	0,02	0,05	W16	1	0,012	0,054	0,01	0,05
W17	1	0,008	0,0334	0,01	0,03	W16	1	0,016	0,054	0,02	0,05
W17	1	0,005	0,034	0,00	0,03	W17	1	0,005	0,074	0,00	0,07
Total	76	2,130	4,9554	0,03	0,07	Total	67	1,781	5,34	0,03	0,08

Test 2

A comparison between the THUs, volume and weight shipped in weeks 11 to 18 (information available in the system for historical data) with the forecast calculated in the proof of concept was done. It was observed a difference between them, especially in terms of THUs and volume (figure 19).

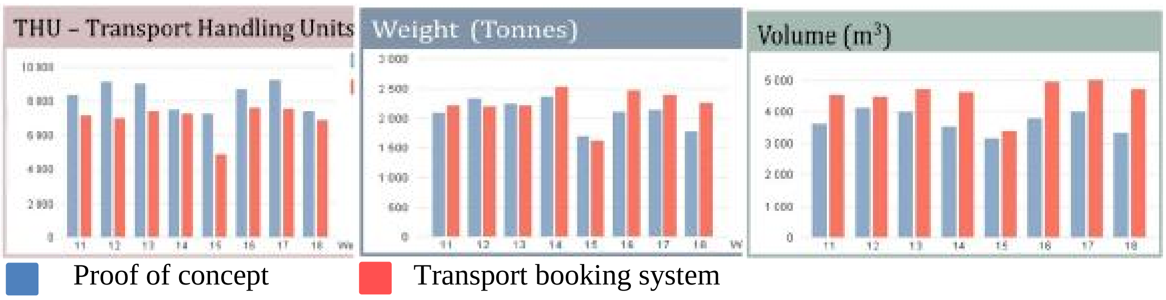


Figure 19 – Results of test 2

7 Analysis and Discussion

7.1 Analysis of the proof of concept

The purpose of this section is to analyze and discuss if the proof of concept developed can provide benefits to the focal company. With this purpose, initially, the results of the proof of concept will be evaluated and the root cause identified; secondly, the barriers identified during the project will be discussed taking as reference the barriers presented by Sanders (2014); finally, the benefits of the proof of concept and its future versions will be evaluated in the form of a roadmap.

7.1.1 Evaluation of the results of the proof of concept

The purpose of this section is to identify the factors that affect the results of the proof of concept presented in the section 6.4.4, as well as identifying the benefits and barriers based on the framework presented in section 3.2.3 and 3.2.4.

For the number of THUs to be shipped, it was observed that in some cases, the quantity of THUs sent was lower or higher than the forecast. After analyzing the shipments, two main causes were identified:

- In some cases, the quantity of part numbers sent was not the quantity in the forecast: more or less pieces were shipped. This is related to the current frozen windows for materials forecast, which varies between 2 to 30 days, translating into possibilities for the supplier and the material controller to change the delivery date or the quantity.
- It is possible that the difference of THUs has been shipped, but it is registered in a different system than the one used for this comparison, as it may have been shipped as a 'premium' delivery.

For the weight and volume per THU shipped, in some cases, the amount shipped matched with the forecast. After an analysis of the quantities shipped, the manual calculation of the volume and weight, as well as a verification of the packaging instructions, it has been concluded that:

- In some cases, the volume or weight entered in the system by the material supplier when booking is incorrect.
- As mentioned in section 6.3.1, some cases with wrong gross weight and volume were identified in the packaging instructions. Nevertheless, there is a possibility of wrong data still remaining in the system, as this is entered manually and can be subject of human error.

7.1.2 Evaluation of barriers

Through the use of the CRISP-DM methodology and the results from the proof of concept, it was possible to identify and evaluate different barriers that would hinder the implementation of the transport demand forecast as a tool to make decisions. The framework presented by Sanders (2014) was used as a way to evaluate the barriers:

Technology

The data necessary to make a transport forecast is in different systems, and it was necessary to use a tool, like Qlikview to merge the data and get value from it. In the case of the proof of concept, it was not a costly process, because of the reduced scope of the project: just one assembly plant. Nevertheless, it is possible that, when developing the proof of concept into further versions, more costly technology is needed in order to manage the big amount of data with the required velocity.

People

When developing the proof of concept, no evidence of barriers related to people, as mentioned by Sanders (2014), was identified: there was no lack of talent within the teams cooperating in the proof of concept, and so far, culture was not identified as a hindering factor for the development or the implementation of the proof of concept. This doesn't necessarily mean that the barriers mentioned by Sanders (2014) are not present, but that, at the proof of concept level this didn't represent an obstacle.

Processes
Through this study, it was possible to identify the lack of integration of the processes and actors related to material transport: the assembly plants communicate a forecast to their material suppliers, nevertheless, the logistics department has no visibility. Even if the actors may identify the impact of their decisions upstream or downstream, they are not able to convert them into tangible impacts. However, it is possible that, developing and implementing the proof of concept will help to connect the processes, increasing the integration of the supply chain.

Data quality
During this study, it was possible to identify a barrier not mentioned by Sanders (2014) in his framework: data quality. As it was shown in previous sections, the results of the proof of concept didn't have the accuracy that was expected, and the root cause was the errors in the information entered in the systems. It has been identified that, in the case of the focal company, this may result from the combination of the three first barriers mentioned by Sanders (2014). For example, the packaging instructions are only used by material suppliers and the assembly plants, therefore, in case it contains an error in the weight or volume, there will be no major impact as the process are not integrated, the systems don't communicate, and nobody else uses the information; as a consequence, there is no urgency to fix the error, as the people may be used to deal with it.

7.1.3 Benefits of the proof of concept

Through the use of the CRISP-DM methodology, it was possible to transform a need from the business, increased visibility, into a tool that is able to translate the material demand into a transport demand for one of the assembly plants. Even though the tool is not able to predict the demand without error, as it has been observed in section 6.4.4, an analysis of the empirical findings combined with the theoretical framework has allowed to identify benefits in case the proof of concept is developed into an implementation state and further 'upgrades' or improvements are done, resulting in a roadmap that explains what additional information would provide what benefits (figure 20).

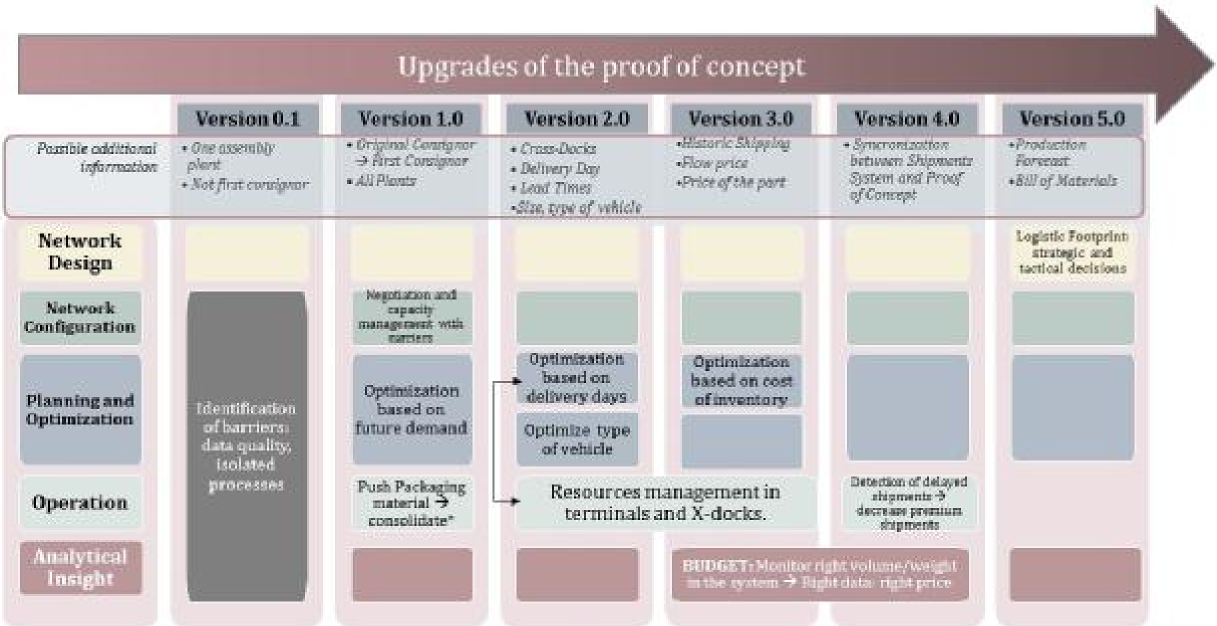


Figure 20 – Future possible development of the proof of concept

Version 0.1
The development of the proof of concept allowed the identification of the barriers previously described: low data quality and isolated processes. This identification, per se, is a benefit as it indicates to the focal company that decisions are being made based on unreliable data: incorrect volumes and weights are taken to project the future and make contracts and capacity estimations; at the same time, they are also the basis for invoice payment: the focal company pays its carriers based on the volume and weight that has been transported. It has been determined that, if the information in the system is incorrect, there can be two consequences: 1) for some shipments, the price that has been paid doesn't reflect the reality of the shipment; 2) the process related to invoice payment will have a high administration cost, as manual checks on volumes may be done by the focal company's employees in order to proceed. The more unreliable the input data is, the more unreliable the output will be.

Version 1.0
As a future development of the proof of concept, it has been considered a scenario where all the material forecasts of all the plants can be processed and translated into a transport forecast. In such case, the focal company would be benefited in different ways: a) in a network configuration level, the purchasing department will be able to provide with forecast volumes to the carriers, instead of historical data, this will allow them to provide more accurate prices per route, as well as foresee if they have a capacity problem for one or several of the routes contracted by the focal company; b) in a planning and optimization level, it will be possible to detect future trends or peaks in the transport demand, allowing the optimization team to establish a direct communication with the carriers and the internal customers (the assembly plants) in order to reach an optimized scenario; c) in the operational level, it will be possible to foresee the demand of packaging material from the material suppliers, as the focal company will be able to identify when the supplier will need more packaging based on the current and forecasted shipments.

Version 2.0
Version 2 of the proof of concept would consider the addition of cross docks, delivery days, lead times and size and type of vehicles used per route. This additional information would allow the optimization team to balance the deliveries per weekday, in a way that the receiver, i.e. the assembly plant, the cross-dock or the terminal, doesn't receive too many goods in one day, which will translate into a more efficient use of the manpower. Additionally, the aggregation of the volumes per supplier and location, will allow the optimization team to identify when it is necessary to move the size or type of vehicle used; for example, if the goods delivered by one supplier are over the capacity of the truck used, then it will be recommended to review if a bigger truck should be used, or if it could move into an FTL scheme.

Version 3.0
A scenario where historic shipping from the transports system, price of the parts and price of the flows are added to the proof of concept was evaluated as well. This additional information would provide two main benefits: the optimization department would be able to balance the shipments based on cost of inventory; it would be possible to make a budget for the inbound logistics costs and identify the deviations from the budget: for example, in case wrong volume is entered in the transport system, it would be possible to detect it and fix it before it generates additional administrative costs or wrong payments.

Version 4.0
The real time synchronization between the proof of concept and the shipments system would allow the focal company to identify as soon as there is a deviation from the forecast: a delay in a shipment or even the possibility to have a premium shipment. This would have an organizational impact, as the focal company doesn't have the capability and flexibility to manage the deviations in a preventive way, i.e. avoid a premium shipment before it happens.

Version 5.0

Finally, in version 5 of the proof of concept information like the production forecast as well as the Bill of materials, BOM, would be added. As a result, it will be possible for the focal company to determine the logistics footprint of their production: what is the logistics cost associated to the manufacturing of certain product, and how this logistics cost is affected if the production varies or moves to a different plant; this would enable the management of the focal company to take tactical and strategic decisions in an easier and more reliable way.

7.2 Big data to support supply chain strategy

Big data can be seen as an important element of the overall strategy of the focal company. By introducing big data solutions as a strategy element, the focal company could materialize several benefits, although there are some barriers that it is imperative to overcome in order to implement the big data strategy. In this section, it will be analyzed big data as an element to support the supply chain strategy, as well as the current barriers that need to be overcome to reach the benefits.

7.2.1 Benefits of the implementation of big data opportunities identified

During the pre-analysis of this study, it was possible to identify fourteen big data opportunities that could support the supply chain strategy of the company in terms of responsiveness and cost. Each of them would provide one or several benefits for the focal company, as it has been described in section 5, and highlighted during the survey conducted (appendix 1). These big data opportunities could be classified according to the framework presented in section 3.2.3, where four main categories of big data benefits were introduced: a) opportunities that predict the future, b) opportunities that increase operational efficiency, c) opportunities that increase understanding and d) opportunities that can lead to new business models. Figure 21 presents the results of such classification.

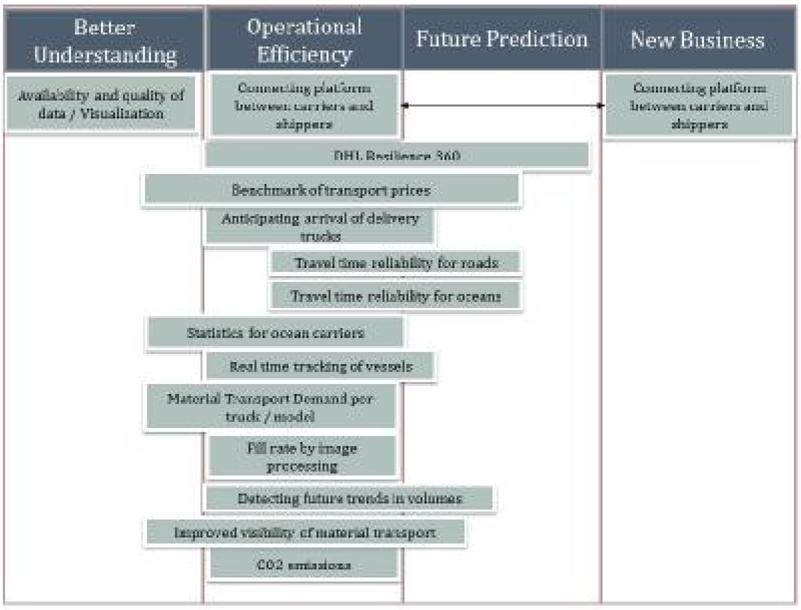


Figure 21 – Fourteen big data opportunities classified

It has been identified that all of the ideas, but availability and quality of data, provide a benefit on operational efficiency, while only one of the big data opportunities would provide new business. Some of the ideas contribute, in higher or lower degree, to other areas like future prediction or better understanding. The result of this analysis is consistent with how the ideas were identified: the main purpose of them is to support the supply chain strategy in terms of responsiveness and cost.

When analyzing the whole supply chain of the focal company, it can be said that the material transport process is located more in the upstream, i.e. there isn't close contact to customers. This can be one of the reasons why the big data opportunities identified don’t provide the bigger benefits into areas like future prediction or new business.

7.2.2 Barriers for implementing a big data strategy

As mentioned previously in this section, these benefits cannot be materialized unless some barriers are overcome as well. These barriers fall into three different categories as presented in section 3.2.4: technology, people and processes.

In terms of technology among the fourteen big data opportunities the main barrier connects to the compatibility issues between the different systems that are used internally in the focal company. For example in order to achieve the implementation of a solution like Improved Visibility of Material Transport Demand, it is imperative for the focal company to find ways in order to build bridges between its internal systems like the system that handles the bookings, the system that handles the packaging instructions, the system the handles the material forecasts and several other internal systems depending on the level of the desired detail. There are other cases like already built solutions, for example like DHL 360 resilience that technology for the implementation of which there is no need for any software development inside the focal company. Other more experimental solutions might require software development to be taken place in order to be integrated in the desired level. Although in all of the cases the compatibility barrier remains the prominent one.

In terms of the people category of barriers two of the three aspects of it could be the case in the focal company: leadership support and culture. In terms of leadership support due to the size of the company and its traditional layout over the years there might be levels in the leadership that do not see a competitive advantage to materialize with the use and processing of data. Thus, it is imperative that the teams responsible for implementing a big data solution have the support from leadership when it comes to their initiative. Regarding the culture aspect, there would be changed needed in the way the employees of the focal company take decisions. There are many cases, like in any big organization, that employees prefer to base their decisions on estimations or personal experience rather than reliable data. For example, some solutions would require the users to actively participate in learning of different big data tools and investing time on utilizing their capabilities. This barrier would be critical in an implementation of any big data solution.

Regarding the final category of barriers, it is considered that the design of the different processes inside the focal company can play an important role in the implementation of a big data strategy. More specifically, a cross functional approach and in most of the cases a cross enterprise approach is essential for the success of the big data strategy. There are examples like the material transport demand per truck/plant or the Improved Visibility of Material Transport Demand solutions that would require cross enterprise approach in order to breach the silos created in the different teams involved. Thus a better communication of processes through the focal company is of great importance for the big data strategy.

8 Conclusions and Recommendations

This research aimed at exploring the utility of big data within the material transport process in terms of complexity, benefits, enablers and barriers. For this purpose two research questions were formulated: how can the material transport process benefit from big data on a strategic, tactical and operational level; and what are the benefits, costs and barriers in implementing big data for the material transport process.

Regarding research question 1, how can the material transport process benefit from big data on a strategic, tactical and operational level, an exploratory phase was carried on, which resulted in the identification of fourteen big data opportunities connected to the two main drivers of the supply chain strategy of the department, cost and responsiveness. Through this report, it was possible to observe that the opportunities identified would provide similar benefits to those highlighted by Accenture (2014) in their survey: faster and more effective reaction time to supply chain issues, greater integration across the supply chain and increase in supply chain efficiency.

Research question 2, what are the benefits, costs and barriers in implementing big data for the material transport process, was approached in two ways: initially, benefits were evaluated for all the big data opportunities identified, as it has been explained in research question 1, while the cost for all the big data opportunities was evaluated in terms of complexity. Secondly, a proof of concept was developed through the use of the CRISP-DM methodology, which allowed identifying the barriers highlighted by Sanders (2014), technology, people and processes, as well as one additional barrier related to quality of the data; and identifying in more details the benefits of the development and implementation of the proof of concept, which could be summarized as: improved volume optimization and savings through the identification of trends in the future; better price negotiation with carriers, as more accurate future volume can be estimated; improved management of human resources in terminals, cross-docks and plants' warehouses; better cost control through a comparison of forecasted volumes and actual volumes shipped; possibility to make a budget of the whole material transport process in line with the production forecast; possibility to determine the logistic footprint of the different products or vehicles of the focal company.

It can be concluded that the inbound logistics process of the focal company would gain benefits from the use of big data tools and methodology, as it was proved during this study.

Based on this study, some recommendations can be given to the focal company. First, to utilize the existing methodology, CRISP-DM, as well as the methodology developed in this study, in other areas of the company, with the aim of identifying other opportunities that can provide high benefits with acceptable complexity. Second, the improvement of the quality of the data both in the packaging instructions, as well as in the transport booking system, would result in higher reliability for the information used to make decisions, i.e. more accuracy when making estimations, negotiating volumes, and defining budget and payment. Third, the use of more advanced analysis tools, like Qlikview, would be recommended in order to make more efficient and deeper analysis, not only in the areas studied in this report, but extended to other areas of the organization that require a more complex data gathering to evaluate situations. Finally, it is highly recommended to develop the proof of concept into a further stage where it can work as a bridge that integrates the manufacturing plants with the logistics department of the company and it will provide the mentioned benefits.

9 References

Accenture, 2014. *Big Data Analytics in Supply Chain: Hype or Here to Stay?.* [Online] Available at: http://www.accenture.com/us-en/Pages/insight-global-operations-megatrends-big-data-analytics.aspx
[Accessed 20 January 2015].

Ansari, A. & Modaress, B., 2010. Challenges of outsourcing logistics to third-party providers. *International Journal of Logistics Systems and Management,* 7(2), p. 198.

Booth, W. C., Colomb, G. G. & Williams, J. M., 2008. *The craft of research.* 3rd red. Chicago: University of Chicago Press.

Bubner, N., Bubner, N., Helbig, R. & Jeske, M., 2014. *Logistics Trend Radar: Delivering insight today. Creating value tomorrow!,* u.o.: DHL Customer Solutions & Innovation.

Chapman, P. o.a., 1999. *CRISP-DM 1.0 - Step-by-step data mining guide,* u.o.: u.n.

Cheong, M. L., 2003. *Logistics Outsourcing and 3PL changes,* Singapore: Nanyang Technological University.

Chopra, S. & Meindl, P., 2013. *Supply chain management: strategy, planning, and operation.* 5th red. London: Pearson.

Christopher, M., 2011. *Logistics & Supply Chain Management.* 4th red. Harlow: Prentice Hall.

CSCMP, 2013. *Glossary of Terms.* [Online] Available at: https://cscmp.org/sites/default/files/user_uploads/resources/downloads/glossary-2013.pdf?utm_source=cscmpsite&utm_medium=clicklinks&utm_content=glossary&utm_campaign=GlossaryPDF
[Accessed 14 May 2015].

Davenport, T. H. & Patil, D. J., 2012. Data Scientist. The Sexiest Job of the 21st Century. *Harvard Business Review*, October, pp. 70-76.

Dhar, V., 2013. Data Science and Prediction. *Communications of the ACM*, December, 56(12), pp. 64-73.

DHL, 2015. *DHL Resilience 360.* [Online] Available at: http://www.dhl.com/content/dam/downloads/g0/logistics/resilience360/dhl_resilience_360_flyer_en.pdf
[Accessed 30 March 2015].

DHL, 2015. *www.dhl-discoverlogistics.com*. [Online] Available at: https://www.dhl-discoverlogistics.com/cms/en/course/services/tts/tts_future.jsp
[Accessed 10 May 2015].

Dumbill, E., 2012. Getting Up to Speed with Big Data. i: *Big Data Now Current Perspectivis from O'Reilly Media.* Sebastopol, CA: O'Reilly Media, Inc..

Gleissner, H. & Femerling, C. J., 2013. *Logistics: Basics -- Exercises -- Case Studies.* Cham: Springer.

Gurbaksh, C., 2013. *2014: The Year Marketer's Big Data Gets Real.* [Online] Available at: http://www.wired.com/2013/12/2014-year-marketers-big-data-gets-real/
[Accessed 30 March 2015].

Handfield, R. B. & Nichols, E. L., 2002. *Supply Chain Redesign: Transforming Supply Chains into Integrated Value Systems.* 1st red. u.o.:Financial Times/ Prentice Hall.

Harrington, L. & Smith, R. H., 2013. *Lean and resilient. The new automotive supply chain.* [Online] Available at: https://s1897772577.t.eloqua.com/e/f2
[Accessed 21 03 2015].

Hendriks, M. P. M., 2009. *Multi-step Optimization of Logistics Networks: Strategic, Tactical and Operational Decisions.* Eindhoven: Technische Universiteit Eindhoven.

Jeske, M., Gruner, M. & Weiss, F., 2013. *Big Data in Logistics: A DHL perspective on how to move beyond the hype,* Troisdorf, Germany: DHL Customer Solutions & Innovation.

Kopalle, P., 2014. *Why Amazon's Anticipatory Shipping Is Pure Genius.* [Online] Available at: http://www.forbes.com/sites/onmarketing/2014/01/28/why-amazons-anticipatory-shipping-is-pure-genius/
[Accessed 09 May 2015].

Kothari, C., 2004. *Research methodology: methods & techniques.* New Delhi: New Age International (P) Ltd.

Kurgan, L. A. & Musilek, P., 2006. A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review,* 21(1), pp. 1-24.

Manyika, J. et al., 2011. *Big data: The next frontier for innovation, competition, and productivity,* s.l.: McKinsey Global Institute.

McAfee, A. & Brynjolfsson, E., 2012. Big Data: The Management Revolution. *Harvard Business Review,* Issue October, pp. 60-68.

McIntire, J., 2014. *Supply chain visibility: from theory to practice.* Surrey: Gower.

Oxford University Press, 2011. *A Dictionary of Media and Communication.* Oxford: Oxford University Press.

Press, G., 2013. *A Very Short History Of Data Science.* [Online] Available at: http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/
[Accessed 18 February 2015].

Qlik, 2015. *Qlik website.* [Online] Available at: http://www.qlik.com/us/company
[Accessed 30 March 2015].

Riege, A., 2003. Validity and reliability tests in case study research: a literature review with "hands-on" applications for each research phase. *Qualitative Market Research: An International Journal,* 6(2), pp. 75-86.

Sanders, N. R., 2014. *Big Data Driven Supply Chain Management.* New Jersey: Pearson Education, Inc..

Schmarzo, B., 2013. *Big Data: Understanding How Data Powers Big Business.* Indianapolis: John Wiley & Sons, Inc..

Schroeck, M., Shockley, R., Romero-Morales, D. & Tufano, P., 2012. *Analytics: The real-world use of big data How innovative enterprises extract value from uncertain data,* New York: IBM Global Services.

Taleb, N. N., 2015. *Silent Risk, Lectures on Probability, Vol. 1.* Draft red. u.o.:Descartes Publishing.

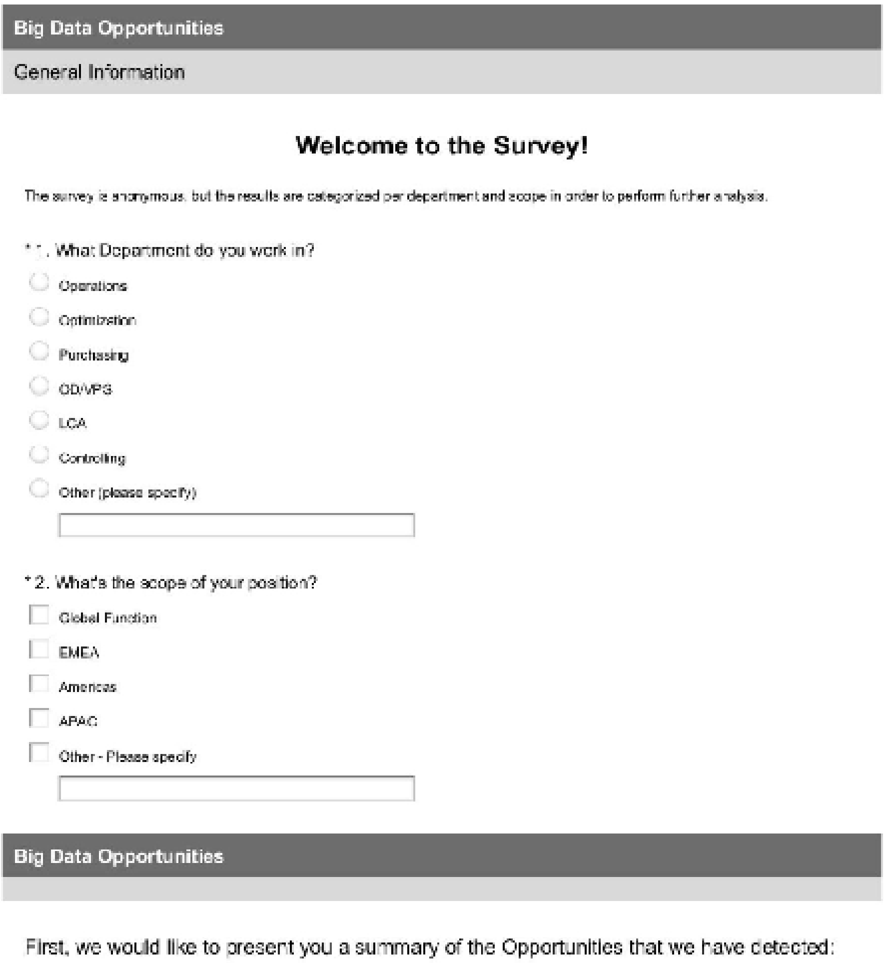
Waller, M. A. & Fawcett, S. E., 2013. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics,* pp. 77-84.

Yin, R. K., 1994. *Case study research: Design and methods.* 2nd red. Newbury, CA: Sage Publications.

Zikopoulos, P. o.a., 2012. *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data.* u.o.:McGraw-Hill.

10 Appendices

10.1 Appendix 1 – Survey and summary of the fourteen big data opportunities


Big Data Opportunities

General Information

Welcome to the Survey!

The survey is anonymous, but the results are categorized per department and scope in order to perform further analysis.

*1. What Department do you work in?

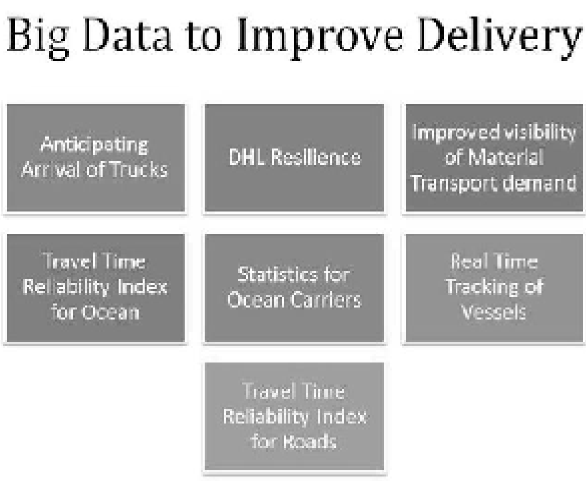
- Operations
- Optimization
- Purchasing
- OD/VPS
- LCA
- Controlling
- Other (please specify)

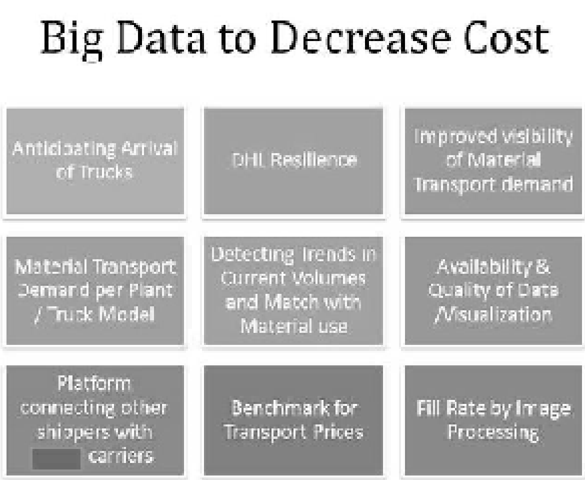
*2. What's the scope of your position?

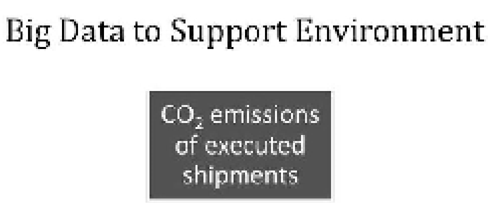
- Global Function
- EMEA
- Americas
- APAC
- Other - Please specify

Big Data Opportunities

First, we would like to present you a summary of the Opportunities that we have detected:


Big Data to Improve Delivery
Anticipating Arrival of Trucks
DHL Resilience
Improved visibility of Material Transport demand
Travel Time Reliability Index for Ocean
Statistics for Ocean Carriers
Real Time Tracking of Vessels
Travel Time Reliability Index for Roads


Big Data to Decrease Cost
Anticipating Arrival of Trucks
DHL Resilience
Improved visibility of Material Transport demand
Material Transport Demand per Plant / Truck Model
Detecting Trends in Current Volumes and Match with Material use
Availability & Quality of Data /Visualization
Platform connecting other shippers with carriers
Benchmark for Transport Prices
Fill Rate by Image Processing


Big Data to Support Environment
CO_2 emissions of executed shipments

Now, we will present, individually, each of the fourteen Opportunities. We ask you to evaluate its benefits based on the next scale:

1 - No Benefit - The solution has no impact on KPIs: Doesn't address any current problem, doesn't improve delivery times, doesn't provide insight for better decision making, or doesn't save money or time.

2 - Low Benefit - The solution has low impact on KPIs: Addresses a sporadic problem with low financial consequences, has small effect on delivery times, provides little improvement on insight for decision making, or saves small amount of money or time.

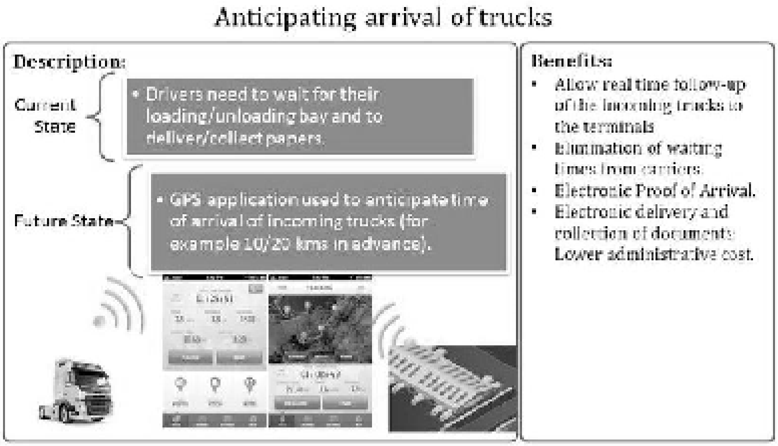
3 - Medium Benefit - The solution has medium impact on KPIs: Addresses a recurrent problem with medium financial consequences, has medium effect on delivery times, saves considerable amount of money or time, or provides considerable improvement on insight for decision making.

4 - High Benefit - The solution has high impact on KPIs: Addresses a frequent problem with high financial consequences, has a major effect on delivery times, reduces an important amount of money or time, or improves the decision making by providing a new perspective.

5 - Very High Benefit - The solution has a very high impact on KPIs: Addresses a critical problem with high financial consequences, will highly improve the delivery times, reduces a big amount of money or time, or improves the decision making by providing a new perspective or time.

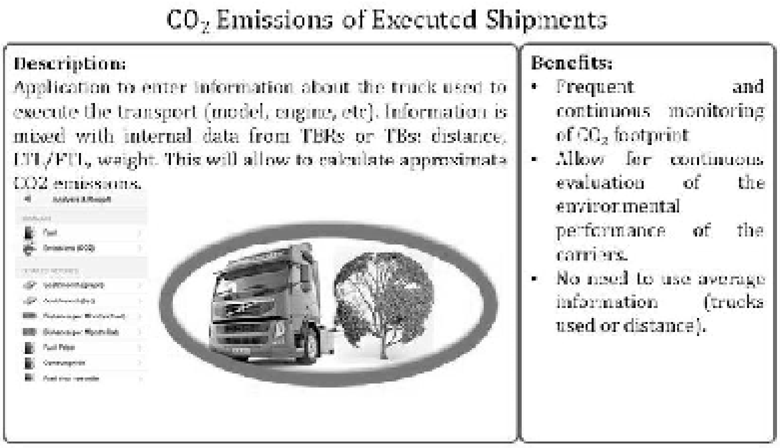
N/A - The opportunity presented is not understood and more information would be needed in order to evaluate it.

In the necessary case, please feel free to add more comments about your evaluation, i.e. why you consider it as a 1, 2, 3, 4 or 5.



* 3. How would you rate the benefit of this solution for the organization?

4. Feel free to provide additional arguments or comments about your answer:



* 5. How would you rate the benefit of this solution for the organization?

6. Feel free to provide additional arguments or comments about your answer:

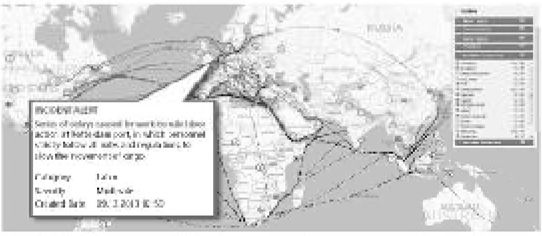
For more information

DHL Resilience 360

Description:

Incident Monitoring: In near real time, tracks relevant risk incidents with potential impact on the end-to-end supply chain; Creates notification alerts; sends immediate feedback on incident status.
Risk Assessment: Exposes vulnerabilities in 20 risk categories; Risk databases cover areas of natural disaster, socio-political and operational risk - scores with ratings from insignificant to extreme

DHL RESILIENCE 360: SUPPLY CHAIN RISK MANAGEMENT



Benefits:

- Minimize effects of an incident through booking of alternative shipments
- Reduce number of premium shipments.
- Better evaluation of the Network Design, as regional or local risks can be identified.

For more information:
http://www.dhl.com/en/logistics/supply_chain_solutions/what_we_do/resilience_360.html

http://www.dhl.com/en/logistics/supply_chain_solutions/what_we_do/resilience_360.htm

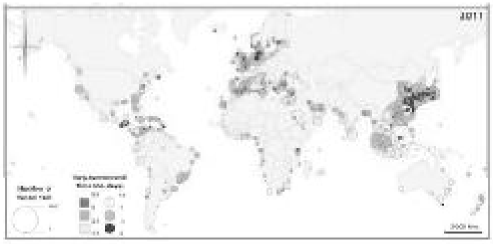
* 7. How would you rate the benefit of this solution for the organization?

8. Feel free to provide additional arguments or comments about your answer:

For more information

Travel Time Reliability Index for Ocean

Description:
Travel Time Reliability is the consistency in travel time for a determined route. For oceans, it considers as well the reliability/performance ports where a vessel stops.



Benefits:
Ability to calculate Reliability Index in travel times would allow:
- Consider more reliable travel times when estimating lead times → impact on inventories and On Time Delivery.
- When configuring the network, it would give insight for accepting/rejecting a route or location based on the Reliability.

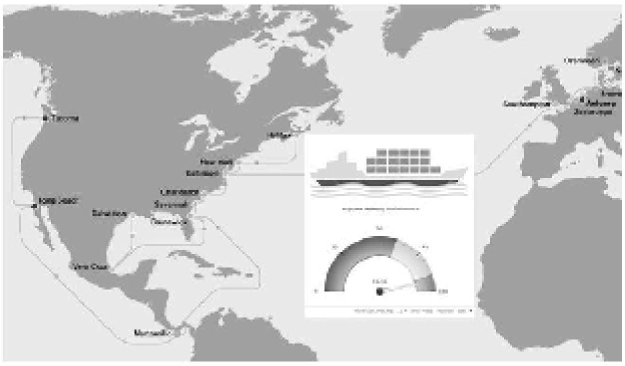
For more information: [illegible]
[illegible]

* 9. How would you rate the benefit of this solution for the organization?

10. Feel free to provide additional arguments or comments about your answer:

Statistics for Ocean Carriers

Description:
External data regarding the performance of Ocean Carriers: On Time Delivery: Actual time vs. Contract Time or Published Ship Schedule.



Benefits:
- Better decision making for supplier selection.

For more information: https://www.youtube.com/watch?v=mrQzwv8wgzA

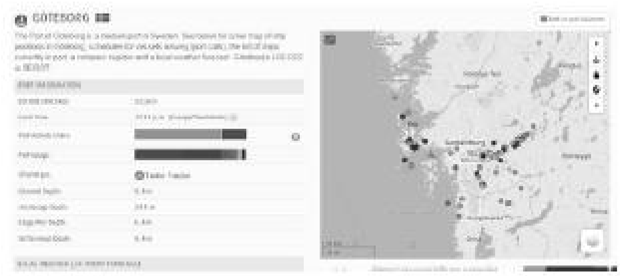
* 11. How would you rate the benefit of this solution for the organization?

12. Feel free to provide additional arguments or comments about your answer:

Example of Vessels Follow-up

Real Time Tracking of Vessels

Description:
Online Services provide information that help shippers to identify schedule changes with impact their shipments. In some cases, they have as well interfaces that deliver live shipping data for software applications.



Benefits:
- Providing vessel tracking would allow to identify if a delay will impact some key delivery.
- Allows to plan and book emergency/alternative shipments, in case a delay is foreseen.

For more information: http://www.fleetmon.com/en/ports/Goteborg_5004

* 13. How would you rate the benefit of this solution for the organization?

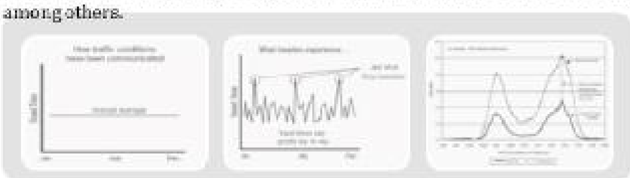
14. Feel free to provide additional arguments or comments about your answer:

For more information

Travel Time Reliability Index for Road

Description:

Travel Time Reliability is the consistency in travel time for a determined route and is affected by congestion, road infrastructure and weather, among others.



Usually

- Travel times are calculated based on average speeds or average traffic conditions.
- They are considered as 'static' through the time.

Reality

There are certain days or times when traffic is higher and therefore, the travel time is longer.

Opportunity

Most effective methods of measuring travel time reliability are 90th or 95th percentile travel times, buffer index, and planning time index.

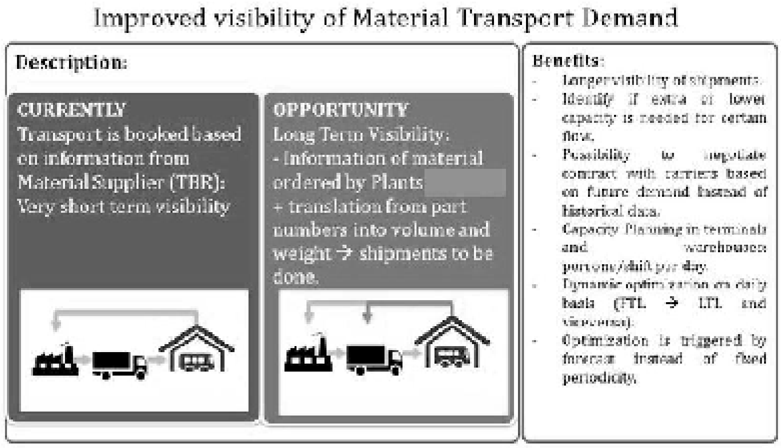
Benefits:

- Determining better Picking Days/Times to be synced with openning hours from Suppliers.
- Consider more reliable lead times when configuring the network → impact on inventories and OTD.
- Avoiding some routes / legs based on low reliability.

For more information: http://www.ops.fhwa.dot.gov/publications/tt reliability/brochure/index.htm

* 15. How would you rate the benefit of this solution for the organization?

16. Feel free to provide additional arguments or comments about your answer:

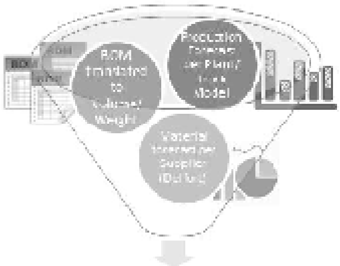


* 17. How would you rate the benefit of this solution for the organization?

18. Feel free to provide additional arguments or comments about your answer:

Material Transport Demand per Plant / Model

Description:
Possibility to combine 1) Production Figures 2) BOM translated into Volume/Weight 3) Material orders per supplier, to create an overall picture of transportation needs per Plant



Shipment Figures per Plant / Model

Benefits:

- Possibility to evaluate the current need of shipments and costs per Plant and Truck.
- Possibility to analyze how a change in the allocation of production will affect the Logistic Network *(Transport capacity and costs)*.

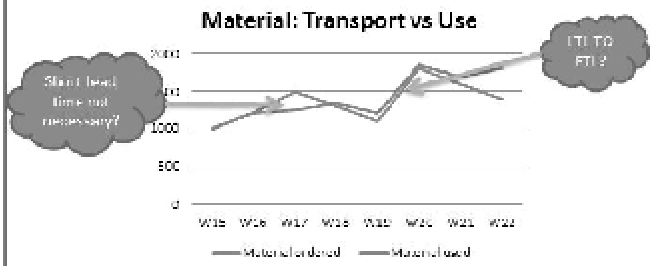
* 19. How would you rate the benefit of this solution for the organization?

20. Feel free to provide additional arguments or comments about your answer:

Detecting Trends in Current Volumes and Match with Material use

Description:

- Monitoring of shipped volumes and detection of trends.
- Matching volumes in transit with actual use of material to adapt lead time.



Benefits:

- Optimization triggered on-demand instead of fixed periodicity.
- Flexible lead times based on demand and use of the materials: if safety stock is high, then we can take longer lead times with lower cost?

* 21. How would you rate the benefit of this solution for the organization?

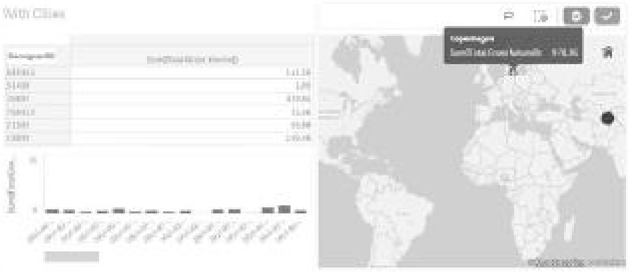
22. Feel free to provide additional arguments or comments about your answer:

For more information

Availability & Quality Data / Visualization

Description:

Big Data Tools that allow to merge and clean data from different sources or databases. Optionally, they can provide improved insight trhough map-style visualizations, volumes trend and origin, among many other options.



Benefits:

- Easier extraction, cleaning and visualization of historical data, allowing personnel to focus on analysis.
- Improved insight that allows better decision making

For more information: www.qlik.com

* 23. How would you rate the benefit of this solution for the organization?

24. Feel free to provide additional arguments or comments about your answer:

Platform connecting other shippers with carriers

Description:

Business Opportunity: Create a regional platform where current transport suppliers, with good delivery performance, offer their services to other companies. Available space (Empty Travel) in LTL or low demand flows can be offered.

Benefits:

- Increase the attractiveness of certain flows: carriers have possibility to offer the space in the return trip.
- Carriers aim at improving performance to be part of selected transport network.

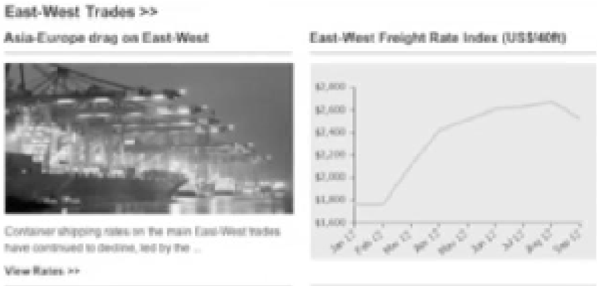
* 25. How would you rate the benefit of this solution for the organization?

26. Feel free to provide additional arguments or comments about your answer:

For more information

Benchmark for Transport Prices

Description:
Use of Big Data Analytics carried out by Market Analysts to benchmark container transport prices before tendering or negotiating. Access to latest news affecting capacity or rates.



Benefits:

- Allow for real-time follow-up of the price levels in the carriers market.
- Providing information when configuring the network or the routes to be used.
- Allows to set target prices before tendering or negotiating a route or a region.

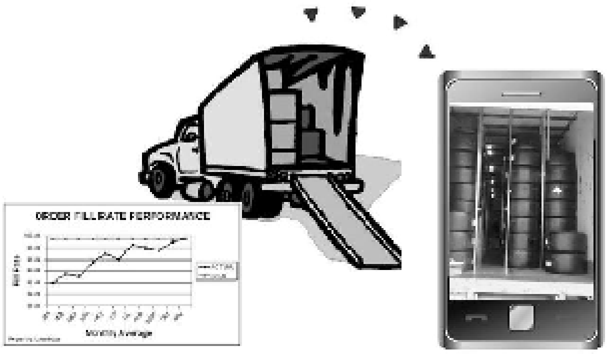
For more information: http://www.drewry.co.uk/publications/view_publication.php?id=350

* 27. How would you rate the benefit of this solution for the organization?

28. Feel free to provide additional arguments or comments about your answer:

Fill Rate by Image Processing

Description:
Use of an application to process pictures taken from the back of the trucks in order to scan the truck and calculate the real filling degree.



Benefits:
- Allow for real and automatic follow-up of the fill rate in trucks.

* 29. How would you rate the benefit of this solution for the organization?

30. Feel free to provide additional arguments or comments about your answer:

Big Data Opportunities

31. s there any other Big Data Opportunity that you have in mind? Please share with us!

32. Would you like to provide any additional comment?

10.2 Appendix 2 – Formulas to construct volume and weight

For complete info = CI

Qty_of_THUs= Ceil((Units / CI – Units per Load),1)

VOLUME=
Volume of pallets
CI - Volume * # THUs
WEIGHT=
Weight of the pallets
CI - Weight * # THUs

For K1, K2, L1, etc = EP1

Qty_of_THUs= Ceil((Units / EP1 - Units per Load),1)

VOLUME=
Volume of the pallets
EP1 - Volume * # THUs
*Ceil (Units / EP1 - Units per Load),1,1)

WEIGHT=
Weight of the pallets
Peso external emballage * # THUs
EP1 - Weight ext emb *Ceil (Units / EP1 - Units per Load),1,1)
+ Weight of the boxes
Peso de la caja* #cajas
EP1 – Weight int emb * Ceil(Units /EP1 - Units per Int emb),
+ Weight of the material
Units * EP1 - Part weight

For K0, L0 =EP2 à 40;k0;780

Qty_of_THUs= Ceil((Units / EP2 - Units per Load),1)

VOLUME=
Volume of the pallets
Volumen external emballage * # THUs
EP2 - Volume ext emb * Ceil((Units / EP2 - Units per Load),1)

+ Volume of the frames
Volume per frame * #frames
Volume per frame * Ceil(# total internal boxes / internal boxes per frame),1
EP2 - Volume per frame * Ceil(Units / EP2 - Units per Int emb),1 / EP2 - Int Emb per frame),1

WEIGHT=
Weight of the pallets
Peso external emballage * # THUs
EP2 - Weight external emb * Ceil(Units / EP2 - Units per Load),1,1

+ Weight of the boxes
Weight of the box * qty of boxes
EP2 – Weight int emb * Ceil(Units / EP2 - Units per Int emb), multiple of EP2 - Int Emb per frame

+ Weight of the material
Units * EP2 - Part weight

10.3 Appendix 3 – Script of data loading for Qlikview

```
SET ThousandSep=' ';
SET DecimalSep=',';
SET MoneyThousandSep='.';
SET MoneyDecimalSep=',';
SET MoneyFormat='#.##0,00 kr;-#.##0,00 kr';
SET TimeFormat='hh:mm:ss';
SET DateFormat='YYYY-MM-DD';
SET TimestampFormat='YYYY-MM-DD hh:mm:ss[.fff]';
SET MonthNames='jan;feb;mar;apr;maj;jun;jul;aug;sep;okt;nov;dec';
SET DayNames='må;ti;on;to;fr;lö;sö';


LOAD [Parma No] as [Supplier No],
     [Part No],
     [Part Desc GPT],
     [Delivery Plant],
     [Unique Number],
     Issue,
     F7,
     [Packaging Type],
     UnitLoad,
     [Gross Weight],
     Volume,
     [I EMB No],
     [Units per I EMB],
     [KOLA Part Weight],
     F15,
     [# of boxes per pallet],
     [Estimated Units per THU],
     F18,
     [CI – Units per Load],
     [CI - Volume],
     [CI - Weight],
     F22,
     [EP1 - Units per Load],
     [EP1 - Volume],
     [EP1 - Weight external emb],
     [EP1 – Weight int emb],
     [EP1 -  Units per Int emb],
     [EP1 - Part weight],
     F29,
     [EP2 - Units per Load],
     [EP2 - Volume ext emb],
     [EP2 - Volume per frame],
     [EP2 -  Units per Int emb],
     [EP2 - Int Emb per frame],
     [EP2 - Weight external emb],
     [EP2 – Weight int emb],
     [EP2 - Part weight],
     F38,
     F39,
     F40,
     F41,
     F42,
     F43,
     F44,
     F45,
     F46,
     F47,
     F48,
     F49,
     [Int_Emb Weight],
     [Int_Emb Volume],
     [Additional Volume per layer],
     [I EMB per THU],
     F54,
```

```
      [Created by],
      [Estimated Volume],
      [Gross Weight1]
FROM
[C:\Projects\Itzel\LOGIC V0.3\Packaging Instructions (GPT).xlsx]
(ooxml, embedded labels);

LOAD [Original Consignor ID] as [Supplier No],
      [OC REF ID],
      [OC NAME],
      [OC CITY],
      [OC COUNTRY],
      [First Consignor ID],
      [FC customer ref],
      Register,
      [FC NAME],
      [FC CITY],
      [FC PROVINCE CODE],
      [FC COUNTRY],
      [FC ZONE2],
      [FC POSTAL CODE],
      [FC ADRESS1],
      [FC ADRESS2],
      [Delivery Plant],
      [Customer Unit ID],
      [CU NAME],
      [CU PROVINCE CODE],
      [CU COUNTRY],
      [CU CITY],
      [CU POSTAL CODE],
      [CUSTOMER UNIT adress1],
      [CUSTOMER UNIT adress2],
      [UNLOAD AT XDOCK],
      [TPT MODE],
      [PLANNING TYPE],
      [LANE CALENDAR XID],
      [PICKUP DAYS],
      [X-REF STATUS],
      [R4 FLOW],
      [DIRECT ROUTING],
      F34
FROM
[C:\Projects\Itzel\LOGIC V0.3\Flows within Europe.xlsx]
(ooxml, embedded labels);



Datatemp:
CrossTable([Week Year], Units, 4)
LOAD [Parma No] as [Supplier No],
      [Part No],
      [Delivery Plant],
      [Unique Number],
      [42100],
      [42107],
      [42114],
      [42121],
      [42128],
      [42135],
      [42142],
      [42149],
      [42156],
      [42163],
      [42170],
      [42177],
      [42184],
      [42191],
      [42198],
```

```
     [42205],
     [42212],
     [42219],
     [42226],
     [42233],
     [42240],
     [42247],
     [42254],
     [42261],
     [42268],
     [42275],
     [42282],
     [42289],
     [42296],
     [42303],
     [42310],
     [42317],
     [42324],
     [42331],
     [42338],
     [42345],
     [42352],
     [42359],
     [42366],
     [42373],
     [42380],
     [42387],
     [42394],
     [42401],
     [42408],
     [42415],
     [42422],
     [42429],
     [42436],
     [42443],
     [42450],
     [42457],
     [42464]
FROM
[C:\Projects\Itzel\LOGIC V0.3\Material Forecast.xlsx]
(ooxml, embedded labels);

STORE Datatemp into Temp.qvd (qvd);
DROP Table Datatemp;

Forecast_of_Material_Deliveries_W15:
LOAD
    Date(Num#([Week Year]),'MM/DD/YYYY') as [Week Year],
    [Supplier No],
    [Part No],
    [Delivery Plant],
    [Unique Number],
            Units
FROM Temp.qvd (qvd);
```