MASTER THESIS

THE DESIGN OF A

COMMON DATA MODEL

FOR GENERIC SYNCHROMODAL CARGO-TRACKING IN LOGISTICS USING WEB SCRAPING AND BIG & OPEN DATA

WOUTER BOL RAAP
University of Twente - Program Business & Information Technology
Faculty of Electrical Engineering, Mathematics And Computer Science

CAPEGROEP shifting borders
UNIVERSITY OF TWENTE.
DINALOG Dutch Institute for Advanced Logistics
CTT INTERMODAAL TRANSPORT
HST Groep
KUEHNE+NAGEL
neele-vat LOGISTICS

May 20, 2016

AUTHOR

Wouter Bol Raap

Study Program	Business & Information Technology Faculty of Electrical Engineering, Mathematics And Computer Science
Student no.	1003739
Email	w.bolraap@student.utwente.nl

GRADUATION COMMITTEE

Dr. Marten van Sinderen

Department	Computer Science
Email	m.j.vansinderen@utwente.nl

Dr. Maria-Eugenia Iacob

Department	Industrial Engineering and Business Information Systems
Email	m.e.iacob@utwente.nl

Sebastian Piest

Company	CAPE Groep
Email	s.piest@capegroep.nl

Preface

With this thesis, my period as a student has come to an end. 7.5 Years ago, when I started with my student life as an Electrical Engineering student, I never thought that I would graduate as a Business & IT student. This study proved to be a perfect fit for me that combines two worlds that are going to be more and more connected in the future. I hope this thesis will be a great start for an even better future.

When I started with this thesis, I expected the opportunity to apply the knowledge I gathered during the years of studying Business & Information Technology at the University of Twente. Besides applying my knowledge, this period turned out to be the opportunity to learn so much more. I have learned much about how the logistic sector works as well as how technology is and can be used to solve problems in the field. Highlights were the several tours at the collaborating companies to see their daily operations.

I am grateful that CAPE Groep offered me the chance to work on my thesis, used their contacts to introduce me to interesting people and made resources available to guide and help me during this period. I would like to thank Dennis Brugging for his help with designing processes and the translation of those processes in an IT solution and Sebastian Piest, who was always available for questions. Also, thank you to everyone at CAPE Groep who were always available and helpful to answer any question I had.

I would like to thank Maria Iacob, Marten van Sinderen and Sebastian Piest for their time and effort. They provided me over the months with lots of feedback on how to improve this thesis and monitored the scientific quality. For all my years as a student and especially the last few months, I would like to thank you my parents, girlfriend, family and friends, for supporting me, for sometimes asking how things were going, and for sometimes not asking how things were going. Finally, I would like to thank Elisa Ostet for designing the cover of this thesis.

Wouter Bol Raap

Abstract

In logistics, questions as “Where is my container?” and “When does my container arrive?” can often not be answered, which restricts the ability of logistics service providers to be synchromodal. Since logistics is complex and multimodal, goods are rarely transported by one carrier and vehicle. To increase efficiency and save costs in the supply chain, both communication between the different parties in the supply chain and the usage of real-time data must be increased. Currently, communication is limited and the usage of real-time data is inefficient. Logistics service providers use real-time data a-modal meaning that real-time data is mainly used to track the progress of a specific part of the shipment. The data is retrieved manually from a number of websites and sharing this data with other actors in the supply chain is limited. This leads to no end-to-end visibility for the whole supply chain. This research proposes a common data model for an integration platform that increases the ability of logistics service providers to be synchromodal and data sharing among the supply chain. The common data model is designed via a bottom-up approach using results of interviews, observations at different logistics service providers, analyzes of open data on websites and the impact of the integration platform on both business processes and the IT architecture. It is validated against their order definitions, industry standards and identified websites and is tested for four weeks in an operational setting at Schiphol. Based on the designed prototype, the integration platform has great potential to enable logistics service providers to be more synchromodal, increase data sharing among the supply chain and save costs.

Management Summary

Context

To increase efficiency and save costs, logistics service providers want to increase the ability to be synchromodal. Synchromodal transport means that a customer agrees on the delivery of products at a specified cost, quality, and within sustainability targets, but gives the logistics service provider the freedom to decide how to deliver, according to those specifications. Knowledge about the status of their shipments is required in order to plan next transportations as efficient as possible. Simple questions as 'where is my container?' or 'when does my container arrive?' cannot be answered automatically. Currently, real-time information is retrieved manually from internet websites at selected times during the day, often once a day. To increase the ability to be synchromodal, real-time information should be updated more frequently during the day. This research is conducted at CAPE Groep in collaboration with logistics service providers HST Groep, Neele-VAT Logistics, Kuehne-Nagel and Container Terminal Twente and is part of the Synchromodal IT research program of the University of Twente.

Objectives

The retrieval of real-time information and the presentation of this information to users will be done by a logistic integration platform. An important part of the integration platform is a common data model which eases communication between different systems, for example, back-office systems of the logistics service provider. It centralizes information and presents it in an unambiguous format to users. The objective of this research is to design a common data model for the integration platform. The main research question that this research answer is:

What is a common data model for logistics for planned and actual information, orders, statuses and disruptions?

Methods

The research starts by performing a literature study to identify the current state of the art. A bottom-up approach is taken to design the common data model. Multiple interviews and observations are held at the collaborating companies to identify the processes to update shipments and the data required for that from both the back-office systems and the websites. Web sites are analyzed to identify what data is openly available. Architectures are designed to define the impact of the integration platform on both the business processes and the IT architecture. The results are combined with the design of the common data model.

The model is validated to check whether the model can be fed with the required information from back-office systems and websites. The validation is done against different order definitions, logistic message standards, the information available on the internet and internal projects from CAPE Groep. A prototype of the logistic platform is built to check the completeness and performance of the common data model in practice and to show the advantages of the integration platform.

Results

According to the literature, planners of logistics service providers require real-time information to decide quickly and reliable, and increase the efficiency of the supply chain. Integration platforms are recognized to have great potential to increase data sharing among the supply chain, centralizing data, presenting data unambiguously and close the gap between IT and business. A common data model is an important part of the integration platform. Literature shows that for designing such model, problems as semantics, schema, data and communication conflicts must be solved.

During the interviews and observations, information is identified, such as data sources, required data for planners, business processes and decision moments. Users use four different types of websites to retrieve the information they need: Carrier, Terminal and Automatic Identification Information (AIS or ADS) data websites. Based on the results of the interviews, architectures are built showing the current situation and the future situation. The impacts of the integration platform are a decrease of the complexity of the track-and-trace process, a decrease of direct communication with customers and carriers, a higher updating frequency and the ability to design different views of information and to do data mining on historical data. Supply chain partners are easily connected to the platform, increasing the communication among the supply chain.

Based on the previous findings, the common data model is designed. Validation shows that most order definitions of logistics service providers and industry standards contain the information to feed the integration platform. The websites contain the data for the real-time information.

A prototype is designed to test and show the common data model with real shipments and data. The integration platform consists of three layers: the application layer, the integration layer and the web scraping layer. The application layer is a Mendix application that contains a user interface and business logic to process the real-time data from the internet. The integration layer is an eMagiz message bus responsible for routing and translating messages. A web scraping layer is a tool that scrapes data from website pages. Cloudscrape (from April 2016 dexi.io) is used as tool in the prototype. A test has been done for four weeks and showed that the integration platform has great potential for both increasing the ability to execute synchromodal shifts and saving costs. Using the prototype planners should be able to make a more efficient planning based on more accurate data in the platform.

Conclusions

The common data model enables logistics service providers built an integration platform that monitors their shipments real-time and presenting the information as unambiguously way as possible. The integration platform increases the ability of logistics service providers to be more synchromodal while reducing costs and increase communication among the supply chain.

Several recommendations for practice are given. First, the prototype should be tested in a real work, with a high amount of shipment to measure its performance and accuracy. Secondly, the scope of the common data model can be extended by including other business processes, such as customs declaration or scraping vehicle planning. Thirdly, the common data model can be extended with more types of disruptions. Currently, one disruption is added with weather information, but more information, such as traffic information or tides and lock times for inland waters can be added. Finally, the platform collects and stores all information. This historical information can be used for data mining purposes to measure carrier’s KPIs or vehicle’s KPIs.

List of Figures

Figure 1: Graphical overview of the Action Design Research (Sein et al., 2011) 6
Figure 2: Difference between multimodal, co-modal and synchromodal (European Container Terminals, 2011) 11
Figure 3: Results of the pilot study of the modes truck, barge and rail (Lucassen & Dogger, 2012) 14
Figure 4: Summary of expected impacts (ALICE WG2, 2014) 16
Figure 5: Meta-model for integration platforms (Oude Weernink, 2015) 17
Figure 6: Levels of interoperability (Bishr, 1998) 21
Figure 7: Scenario mentioned by Böhmer et al. (Böhmer et al., 2015) 21
Figure 8: Data modelling granularity levels (Lampathaki et al., 2009) 23
Figure 9: Transportation messages in the logistic interoperability model (GS1, 2007) 26
Figure 10: An example mapping of two definitions 27
Figure 11: Unstructured, semi-structured and structured data 29
Figure 12: Screenshot of the interface of CloudScrape 31
Figure 13: Example scenario of a shipment 35
Figure 14: General track and trace process for deep sea 38
Figure 15: Track and trace process for barge 39
Figure 16: General track and trace process for air 40
Figure 17: An example of a supply chain 45
Figure 18: Current architecture 47
Figure 19: Business layer for the current architecture 48
Figure 20: Application layer of the current architecture 49
Figure 21: Technology layer for the current architecture 49
Figure 22: Target architecture 51
Figure 23: Business layer for the target architecture 52
Figure 24: Application layer for the target architecture 52
Figure 25: Technology layer for the target architecture 53
Figure 26: Gap analysis 55
Figure 27: Overview of the CDM design process 57
Figure 28: The data that the CDM consists of 58
Figure 29: The designed common data model presented in Mendix 64
Figure 30: Static order data of the CDM 66
Figure 31: Semi real-time order data of the CDM 67
Figure 32: Real-time data of the CDM including the Trip entity 67
Figure 33: Conceptual design of the integration platform 81
Figure 34: The architecture of the prototype 83
Figure 35: The application overview of the prototype 84
Figure 36: General updating process in Mendix 87
Figure 37: The technology overview of the prototype 89
Figure 38: eMagiz bus by Veldhuis (Veldhuis, 2015) 90
Figure 39: Designed eMagiz bus for the prototype 90

Figure 40: Generic request to Cloudscrape 90
Figure 41: Responses type and the information expected 91
Figure 42: The application and technology overview of the prototype 92
Figure 43: Example Cloudscrape robot 93
Figure 44: Homepage of the user 95
Figure 45: Data types 104
Figure 46: Designed common data model 105
Figure 47: Conceptual design of the prototype 106
Figure 48: Step-for-step process at HST Groep for incoming ships for import seafreight xiv
Figure 49: Step-for-step process for outgoing ships for import seafreight xv
Figure 50: Step-for-step process for outgoing ships from Rotterdam for export seafreight xvi
Figure 51: Step-for-step process for arriving ships for export seafreight xvii
Figure 52: Step-for-step process for airfreight xviii
Figure 53: Import Process at Neele-Vat xix
Figure 54: Export process for the periodically process at Neele-Vat xx
Figure 55: Import process at CTT for barge xxi
Figure 56: The export process for barge at CTT xxii
Figure 57: Current architecture of the logistics provider in the supply chian xxiii
Figure 58: Target architecture of the logistics provider in the supply chain xxiv
Figure 59: Gap analysis of the current and future architecture xxv
Figure 60: Shipment overview xxvi
Figure 61: Leg specific Information xxvi
Figure 62: Message overview xxvii
Figure 63: Settings for updating shipments xxvii
Figure 64: Traject information with weather information xxvii

List of Tables

Table 1: Which research question is treated in which chapter? 8
Table 2: Papers used in Section 2.3 12
Table 3: Papers used in Section 2.4 18
Table 4: Papers used in Section 2.5. Unstructured Data 28
Table 5: Definitions used in Section 3 and up 34
Table 6: List of interviewees 36
Table 7: Data retrieved from website types for Deep Sea 38
Table 8: Data needed from dossier or back office system for Deep Sea 39
Table 9: Data retrieved from web sites type for Barge 40
Table 10: Data needed from dossier or back-office system for Barge 40
Table 11: Data retrieved from web sites type for Air 41
Table 12: Data needed from dossier or back-office for Air 41
Table 13: Sources per website type 42
Table 14: Found sources 42
Table 15: The identified data that is necessary for the integration platform 44
Table 16: Data necessary for modalities Sea and Barge 60
Table 17: Data necessary for modalities Sea, Barge and Air 61
Table 18: Data necessary for modalities Sea, Barge and Air with additional fields 63
Table 19: Description of the different entities of the CDM 65
Table 20: Websites analyzed for the website approach 70
Table 21: Mapping of the Order Definition of HST Groep 72
Table 22: Mapping of the Order Definition of Neele-Vat 73
Table 23: Mapping of the Order Definition of EDIFACT IFCSUM 74
Table 24: Mapping of the Order Definition of EDIFACT IFTMIN 75
Table 25: Mapping of the Order Definition of GS1 Standard Transport Instruction 76
Table 26: Mapping of websites for Deep Sea and Barge 77
Table 27: Mapping of websites for Air 78
Table 28: Static Data Mapping for Seacon Project 79
Table 29: Semi Real-Time and Real-Time Data Mapping for Seacon Project 79
Table 30: Test run results 95
Table 31: List of Interviewees 96
Table 32: Identified additional websites on the World Wide Web 100
Table 33: Identified websites 101
Table 34: Information necessary to update shipments for all modalities 102
Table 35: Applicability of the definitions, websites and projects mapped 106
Table 36: Information used at HST Groep by planners at Import over Sea department xiv
Table 37: Information needed for Import over sea for outgoing ships xv
Table 38: Information needed for the Export Department for outgoing ships xvi
Table 39: Information needed for the Export Department for incoming ships: xvii
Table 40: Information needed by Airfreight planners xviii

Table 41: The data needed for the Import Process xix
Table 42: Data needed for the Export Process at Neele-Vat xx
Table 43: Information needed for Barge import process at CTT xxi
Table 44: Data used for the Export process for Barge at CTT xxii

List of Abbreviations

aPaaS	application Platform as a Service
API	Application Programming Interface
AWB	Air WayBill
B/L	Bill of Lading
CDM	Common Data Model
CSS	Cascading Style Sheets
CTT	Container Terminal Twente
EDI	Electronic Data Interchange
ETA	Estimated Time of Arrival
ETD	Estimated Time of Departue
GHG	Green House Gas
GPS	Global Positioning System
GUI	Graphical User Interface
HTML	HyperText Markup Language
iPaaS	Integration Platform as a Service
IT	Information Technology
LSP	Logistics Service Provider
SCAC	Standard Carrier Alpha Code
TMS	Transport Management System
UML	Unified Modeling Language
XML	EXtensible Markup Language
XSD	XML Schema Definition
XSLT	Extensible Stylesheet Language Transformations

Contents

Chapter 1: Introduction 1
1.1 Introduction 1
1.2 Context 1
1.3 Research Motivation 3
1.4 Problem Statement and Objective 4
1.5 Methodology 5
1.6 Scope 7
1.7 Structure of this report 7
Chapter 2 Literature Review 9
2.1 Introduction 9
2.2 Methodology 9
2.3 Synchromodality 10
2.4 The Common Data Model 18
2.5 Big and Open Data 28
2.6 Conclusion 31
Chapter 3 Interviews, Data and Processes 34
3.1. Introduction 34
3.2. Interview set-up 35
3.3. General Findings 36
3.4. Processes 37
3.5 Websites and Data Representations 41
3.6. Conclusion 43
Chapter 4 The Architecture 45
4.1 Introducing the supply chain 45
4.2 Modelling the architecture 45
4.3 The current architecture 46
4.4 The target architecture 50
4.5 Differences 53
4.6 Conclusion 56
Chapter 5 Design the Common Data Model for Synchromodal Logistics 57

5.1 Introduction 57
5.2 What data is covered in the common data model? 58
5.3. The Common Data Model 59
5.4 Design of the CDM 64
5.5 Conclusion 68
Chapter 6 Validation of the Common Data Model 69
6.1 Introduction 69
6.2 Validation Methods 69
6.3 Validation 71
6.4 Conclusion 80
Chapter 7 Prototyping the CDM into an Integration Platform 81
7.1 Introduction 81
7.2. Architecture of the prototype 82
7.3. Mendix Application 83
7.4. eMagiz 88
7.5. Cloudscrape 92
7.6. Performance of the Prototype 94
7.7. Validation of the Prototype 96
7.8 Conclusion 98
Chapter 8 Conclusion 99
8.1 How to build an Integration Platform? 99
8.2 Data Sources 100
8.3 Data Representations 101
8.4 Data storage and a more efficient supply chain 102
8.5 An Architectural View 103
8.6 The Common Data Model 103
8.7 Development and Testing of the Prototype 106
8.7 Conclusion of this Research 107
Chapter 9 Recommendations 110
9.1 Recommendations for Science 110
9.2 Recommendations for Practice 110
References 112

Appendix A: Interview Set-up and Interview Summary for Process Identification i
Appendix B: Identification of Current Track and Trace Processes...xiii
Appendix C: Architectures xxiii
Appendix D: Screenshots of the prototype..xxvi
Appendix E: Validation and Evaluation Interview Set-up and Interviews xxviii

Chapter 1: Introduction

Logistics is one of the world's largest service sectors. Millions of containers are transported all over the world to deliver products customers need and want. Over the years, customers have become more and more demanding due to the rapid development of technology. Customers demand better service and more accurate and faster delivery. Transportation companies and logistics service providers have difficulties coping with these demands. Synchromodality offers the ability to plan the transportation as optimal, flexible and sustainable as possible using different modes of transportation and enables transportation companies and logistics service providers to cope with customer's demands. The chapter starts with introducing the context of the thesis and introducing the collaborating companies. Section 1.2 shows the research motivation. The research objective and questions are discussed in section 1.3. The scope of the project is set in section 1.4. Section 1.5 explains the methodology of the thesis. The structure of the thesis is explained in section 1.6. Section 1.7 concludes the chapter.

1.1 Introduction

This first chapter introduces the thesis by giving an overview of the context of the problem and the problem different companies have with tracking their shipments and vehicles. Currently, companies track their shipments manually by tracking the vehicle that carriers the shipment. The objective of this thesis is to design the common data model for planned and actual information, orders and status/disruption so that shipments over different modalities can unambiguously and can be compared. Such a common data model stimulates the creation of an integration platform that automatically can track shipments and data sharing among the supply chain. Furthermore, the scope and the structure of the thesis is defined.

1.2 Context

Logistics Service Providers (LSP), from here logistics provider or LSP, can transport freight via motorized transport over different modes (modalities): water (sea and barge), air, road and rail (Runhaar, 2001). For short distances, transport using one modality is common, but for larger distances, reaching inland customers or even reducing Green House Gas (GHG) emissions it is common to use at least two different modes (co-modal) (Steadieseifi, Dellaert, Nuijten, Van Woensel, & Raoufi, 2014). When freight is moved in the same loading unit, for example, a container, using two or more modes of transport without handling the goods themselves when changing modes, this is called multimodal transport (Forum, 2010). Multimodal transport is defined as "the carriage of goods by at least two different modes of transport on the basis of a multimodal transport contract from a place in one country at which the goods are taken in charge by the multimodal transport operator to a place designated for delivery situated in a different country" (Nations, 1980).

With multimodal transport, the planning of the transport from A to B is fixed with locations to change mode or vehicle (Steadieseifi et al., 2014). It is very hard for planners to change schedule and planning if delay occurs along the way. As an example, when problems arise with a ship that holds a customer's container and these problems cause delays, it is possible that the container misses the changeover. If this happens, logistics providers will reschedule, when necessary, while risking higher costs. Synchromodality enables LSPs to use intermodal planning with the possibility of real-time switching between the modes

(Tan, Klievink, Overbeek, & Heijmann, 2011). This makes LSPs more flexible while reducing costs, the existing infrastructure is utilized better and the GHG emissions reduce (Egberink, 2015).

Synchromodality is described as transport where a customer agrees with a logistics provider on the delivery of products at a specified cost, quality, and within sustainability targets, but gives the logistics provider the freedom to decide how to deliver according to those specifications (van der Burgh, 2012).

This research is part of the Synchromodal IT research program of the University of Twente. The goal of the Synchromodal IT research program is to create an integrated synchromodal logistical network to increase the efficiency, reliability and sustainability of Logistics Providers and stimulate a mental shift to let consignees and logistics companies adopt the synchromodal logistics concept (University of Twente, 2016). The program is a unique combination of real-time big data, planning and serious gaming and architecture services.

1.2.1. CAPE Groep

Cape Groep B.V. is a company located In Enschede, the Netherlands and specializes in integrating IT solutions and participates in the Synchromodal IT research program. One of their specialties is supply chain solutions, making sure that these solutions give full control and that communication among various systems used is optimized by matching the in- and outputs of these systems.

1.2.2. Neele-Vat Logistics

Neele-Vat Logistics is a company based in Rotterdam, the Netherlands. It is one of the larger logistics service provider in the Rotterdam region with 10 offices all around the Netherlands and 4 dispersed over Europe (Moscow, Vantaa, Oradea and Prague). Around 500 people work at Neele-Vat and it has a yearly turnover of €230 million (Logistics, 2015). Besides European road transport, they offer physical distribution via air and sea transport and storage of goods. In the top 100 logistics providers of 2015 Neele-Vat Logistics has the 39th place (Logistiek Krant, 2015). During the research, Neele-Vat arranges interview and observation possibilities.

1.2.3. HST Groep

HST Groep is an all-round logistics service provider, based in Enschede, the Netherlands (HST Groep, 2015). HST was found in 1978 when three regional transport companies joined forces. HST has all kinds of disciplines. The service offerings include sea freight, air freight, (inter)national road transport, warehousing and e-fulfilment. Over the years, HST became a (medium sized) international transport company that provides a large range of logistic services. In the top 100 logistics providers of 2015 HST Groep has the 71st place (Logistiek Krant, 2015). During the research, HST Groep arranges interview, observation and validation possibilities.

1.2.4. Kuehne+Nagel

Kuehne+Nagel is a worldwide logistics provider offering transportation over deep sea, air and land. Currently, it has more than 1000 offices worldwide offerings their services in more than 100 countries. It is the number one global sea freight forwarder, number two in global air cargo forwarder and belongs to the top 3 in transportation overland. In the top 100 logistics providers of 2015 in the Netherlands, Kuehne+Nagel has the 3rd place (Krant, 2015). During the research, Kuehne+Nagel arranges validation possibilities.

1.3 Research Motivation

The creation of an optimal, flexible and sustainable usage of different modes of transportation within a network is complex. Currently, logistic companies search online for information about their shipments. Information such as expected arrival time (ETA), the current location of the vehicle and the status of the container is important for the planning of further transportation of the goods and to have knowledge about the progress of the shipment. The value of the ETA is currently determined by the planner by checking the location of a vehicle at various moments during the trip or by checking the ETA at the arriving port and comparing these locations and ETA to shipping schedules, which are collected from the transportation company (Veldhuis, 2015). Data is manually retrieved provided by a number of websites. However, within a growing industry, this costs an increasing amount of time, demanding automated processes to take over in the future.

A solution is a logistic integration platform that integrates customers, partners and information sources and automates the retrieval of the relevant data and makes it available for end users (Oude Weernink, 2015; Toth & Liebler, 2013; Zaiat, Rossetti, & Coelho, 2014). Oude Weernink (2015) designed a meta-model for a logistic integration platform which considers all aspects that are necessary for an integration platform. Oude Weernink also stresses the importance of a common data model (CDM). Each data source or back-end system has their own data formats, protocols and encodings which make communication between each system difficult. A common data model is a standardized definition of how system solutions and technologies represent resources and their relationships (Singh & van Sinderen, 2015). Different challenges of interoperability arise when creating a CDM. Singh and van Sinderen noticed these interoperability challenges in synchromodal logistics.

Data needed by planners, provided by different websites or systems, are available to use but is presented either unstructured or structured in the context of Synchromodality. The data are stored and presented differently at every website. To increase the ability of synchromodal modality shifts, network data of all existing modalities (sea, air, train, barge and road) should be available and actual in a controlled environment (Zaiat et al., 2014).

To store and present all this data for different modalities, which each their own websites, a common data model is necessary. Currently, there is no common data model, willingness to share data is missing and IT maturity of supply chain partners is very diverse. Therefore, it is hard to create an end to end supply chain visibility and synchromodal decision support.

1.3.1 Scientific relevance

This research contributes to multiple aspects to the scientific field. The first contribution is the designed data model that combines and stores the different structured data from different transportation modes, providing the possibility of data mining and analysis. Secondly, the insights of how the supply chain changes when introducing a service using the designed CDM that improves synchromodality.

1.3.2 Relevance for practice

CAPE Groep participates in the Synchromodality program because of its potential. They are interested in the data model and the prototype that automates the search and retrieve of open data on websites. If successful, the project can be expanded to data mining and analysis and could be a first step to calculating actual ETAs. Further developed, it can be a new product that can be sold to different partners in the logistics sector.

1.4 Problem Statement and Objective

Currently, no common data model exists that enables logistics providers to build a logistic integration platform to increase the ability to be synchromodal and communication over the supply chain. Planners need real-time information to make more accurate and reliable decisions. Currently, this information is manually retrieved by accessing different websites (Veldhuis, 2015).

Based on this problem a research goal can be formulated. The goal of this research is to improve the ability to execute synchromodal modality shifts by designing a common data model that covers planned and real-time information about the status of orders. In a logistic platform, this information can be automatically retrieved and has the potential to increase help logistics providers (Oude Weernink, 2015; Zaiat et al., 2014). A common data model is required to be able to translate messages between different end-systems (Oude Weernink, 2015; Zhang, 1994). It will remove conflicts within the data, harmonizes data and present it to users unambiguously (Zhang, 1994). An integration platform will increase data sharing among the supply chain since supply chain partners can be connected easier with each other by connecting to the platform.

This research goal can be translated to the following research question:

RQ: What is a common data model for logistics for planned and actual information, orders, statuses and disruptions?

A number of sub-questions will help to answer the research question. The research starts with getting insights in the current state of the art by performing a literature study. Firstly, the features of synchromodality and its differences with other types of transportation need to be discussed. Secondly, a common data model implies that different end systems need to be integrated via an integration platform. In the literature will be researched why logistics need such integration platform and what is known about integration platforms in logistics. Thirdly, to design a common data model, a method must be chosen in order to design the common data model correctly. At last, the World Wide Web contains a lot of data in different notations and representations. Literature will be studied to find how all this data can be presented unambiguously to the user. This leads to the following sub-questions:

SQ1a: What is the difference of synchromodality in comparison with other types of transportation?

SQ1b: Why do logistics providers need an integration platform?

SQ1c: How is a common data model for logistics designed?

SQ1d: How can big and open data dealt with on the World Wide Web?

Interviews and observations with the collaborating logistics providers must be held to identify the processes, sources and data. Analysis of the sources must be done to identify how data is presented on the websites. Literature and the analysis combined to answer the following sub-questions.

SQ2a: What data sources are used by Logistics Service Providers?

SQ2b: Which data sources are available?

SQ3: What are the similarities and differences between the different representations of the retrieved data?

SQ4: What historical and current data should be stored and shared with the rest of the supply chain in order to make it more efficient?

An integration platform narrows the gap between IT and business (Toth & Liebler, 2013). The impact of the integration platform must be identified and is identified by designing architectures of the current and target situation. An analysis of the differences between the architectures shows the impact.

SQ5: What does the architecture of a supply chain look like from a synchromodal perspective?

When these five sub-questions have been answered, all the information is available to design the data models for planned and actual information, orders and status/disruptions.

SQ6: How to design a common data model for planned and actual information, orders and status/disruptions?

To evaluate the common data model, the model is tested by designing an integration platform that implements the model. This service will be validated and evaluated with the companies that collaborate in this research.

SQ7: How to design a prototype implementing the common data model?

Last, but most certainly not least, this prototype has to be tested effectively, to check whether it provides the functionalities required. Simply translated into a question:

SQ8: Does the prototype have positive effects on the ability to execute synchromodal modality shifts?

With the answer to this final sub-questions, the main research question is completely answered. In the first sub-questions is shown what data is required and from what sources this information can be retrieved. The impact of the integration platform is identified. Based on all these results is shown how to build a common data model and validated that this common data model includes the information that is required to store real-time and planned information and has positive effects on the ability to execute synchromodal modality shifts.

1.5 Methodology

This research follows the Action Design (ADR) methodology of Sein et al. (Sein, Henfridsson, Rossi, & Lindgren, 2011). A graphical overview is shown in figure 1. This methodology has four steps:

1. **Problem Formulation:** In this stage research opportunities are identified and conceptualized and the problem is formulated. The problem can be either practice inspired or theory-ingrained. This research is a practice-inspired research since logistics providers are actively looking for a solution to their problem.
2. **Building, Intervention and Evaluation:** The second stage used the formulated problem and the theory to generate an initial design of the IT artifact. In an iterative process, the IT artifact is built, intervened in the organization and evaluated. Based on the evaluation the IT artifact is further designed.

   This process has two end points. Firstly, if the researcher takes an IT-dominant Approach, the result is an innovative technological design of the IT artifact. Secondly, if the researchers take an Organization-Dominant approach, the result is where the IT artifact causes innovation in the organization.

   This research takes an Organization-Dominant approach. The IT artifact is not innovative within the scientific world, but its applicability in the logistic sector will innovate organizations.
3. **Reflection and learning:** This stage reflects the results of the IT artifact and generalizes the results.
4. **Formalization of Learning:** The objective of the last stage is to formalize the learning of the results.

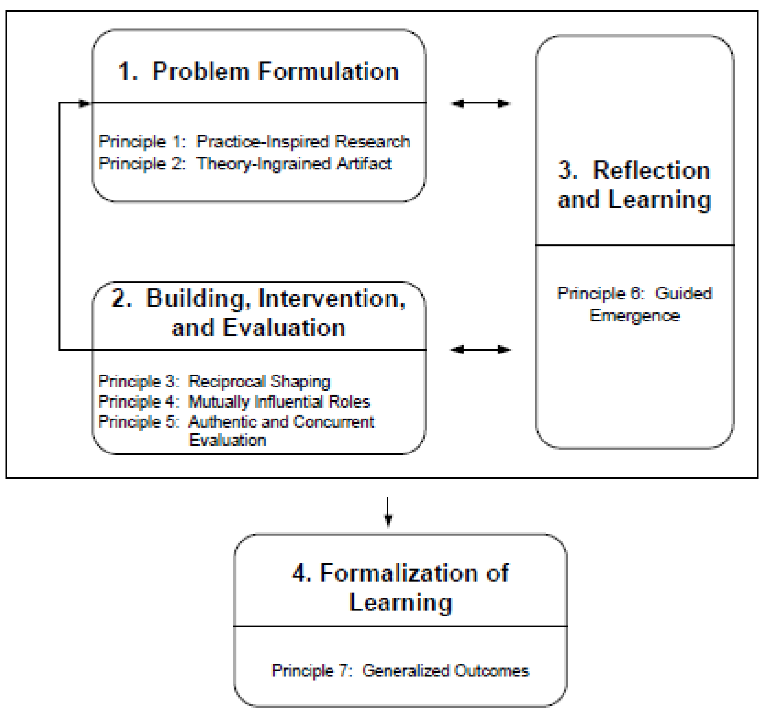

Figure 1: Graphical overview of the Action Design Research (Sein et al., 2011)

1.6 Scope

As described in the previous section, the research goal is to design a common data model. This common data model improves interoperability between systems that use different standards, formats or protocols. Due to limited time and resources available for this research is decided to limit the modalities supported in the common data model to Deep Sea, Air and Barge. For the other modalities Rail and Road, companies have shown none to very little interest for automatically track and tracing. In addition, it is hard to retrieve open data for those modalities. Because of this, these modalities are excluded from the scope of this research. Due to limited time and resources, the scope of the integration platform is limited to it basic functionalities. The integration platform includes the retrieval of data from websites, the integration bus and a portal that can be used by the logistics provider and its supply chain partners to create visibility of the data.

1.7 Structure of this report

The first chapter gives the introduction of this thesis. It covers the initial first step of the ADR and describes the problem identification and motivation from a practical perspective. The second step of the ADR method includes chapters 2 till 7. The second chapter lists several findings based upon research publications and the third chapter lists findings based upon interviews and/or observations. The third chapter shows the results of interviews and observation done at Neele-Vat, HST Groep and CTT. The fourth chapter elaborates about the architectures of the current and target situation and the impact of the integration platform on the logistics provider. The fifth chapter designs the common data model based on the findings of previous chapters. The sixth chapter validates the designed common data model. The seventh chapter designs a prototype of the integration platform implementing the designed common data model. Chapters 6 and 7 include the “Reflection and learning” as well because of the iterative process in which the artifact is reviewed and learnings are taken into account during the next building phase. The eight chapter generalizes and concludes the findings. This corresponds to the fourth step of the ADR methodology. The last chapter suggests possible recommendations for future work as well as recommendations for future research. Table 1 shows what research questions are answered in what chapters.

Table 1: Which research question is treated in which chapter?

Chapter	Research Question
2	SQ1a: What is the difference of synchromodality in comparison with other types of transportation? SQ1b: Why do logistics providers need an integration platform? SQ1c: How is a common data model for logistics designed? SQ1d: How can big and open data dealt with on the World Wide Web?
3	SQ2a: What data sources are used by Logistics Service Providers? SQ2b: Which data sources are available? SQ3: What are the similarities and differences between the different representations of the retrieved data? SQ4: What historical and current data should be stored and shared with the rest of the supply chain in order to make it more efficient?
4	SQ5: What does the architecture of a supply chain look like from a synchromodal perspective?
5	SQ5: What does the architecture of a supply chain look like from a synchromodal perspective?
6	SQ6: How to design a common data model for planned and actual information, orders and status/disruptions?
7	SQ7: How to design a prototype implementing the common data model? SQ8: Does the prototype have positive effects on the ability to execute synchromodal modality shifts?

Chapter 2 Literature Review

This chapter lists several findings based upon research publications and will answer sub-questions 1a till 1d. A structured literature review is performed to identify the concepts related to Synchromodality. By following a structured approach the quality of the work can be safeguarded. This chapter is part of the second step of the ADR method.

Section 2.1 introduces the concepts discussed in this chapter and relates the different concepts. In section 2.2 the methodology will be explained which is used for all subjects in this chapter. Section 2.3 gives a more theoretical context about synchromodality and explains the need for a logistics integration platform. Central in this thesis is the Common Data model. Section 2.4 elaborates what a Common Data Model is and how it is designed. Section 2.5 explains how we can structure and retrieve data. Section 2.6 concludes the findings.

2.1 Introduction

The methodology of the literature review is discussed in section 2.2. Section 2.3 gives more insight into the synchromodal perspective. It explains why the logistics sector has a need for synchromodality, what currently the state of the art is and it explains the need for a logistic integration platform. Concerning the logistic integration platform, there will be explained what research is done on this topic, what open topics are and how to build a logistic integration platform.

The common data model is the center of the integration bus. The bus routes messages to the correct destination. All incoming messages are transformed into a CDM message and all outgoing messages are transformed from a CDM message to the message that is readable and accepted by the receiving system. Section 2.4 elaborates what a CDM message is, methods to design a CDM and what is in the literature about others reference models. A CDM is meant to structure unstructured data. In this context, the unstructured data is dispersed over different systems and all over the internet. Users need structured information in order to efficiently use it. In section 2.5 more information is given about big and open data on the World Wide Web.

2.2 Methodology

For the literature review we follow the method by Tamm et al. (Tamm, Seddon, Shanks, & Reynolds, 2011) since it includes both a systematic and an exploratory review and is recognized by experts as a good method for performing a literature review. First, a systematic review approach is used to identify publications on the specific subject. Secondly, an exploratory approach was used to find additional publications since a systematic review approach might exclude some useful and relevant publications. Both approaches are elaborated below.

2.2.1 Systematic Review

The purpose of the systematic review is to ensure a comprehensive coverage of interesting papers on the topic of synchromodality. To identify relevant literature, two online databases are used: Scopus and Google Scholar. Scopus is chosen for its coverage of academic journals and Google Scholar for its coverage of both journals and books.

First of all, search terms are used to identify the first batch of papers. The search terms are based on the sub-questions the section tries to answer. After removing duplicates and publications that were not available to read and download a selection of papers will remain. These papers are filtered by analyzing the abstracts, the key themes and in what context the subject is used. This results in papers that are all available to read in full text and applicable to the research. To use of the papers efficiently, the papers are categorized by their content and their use. Papers categorized in a subject that is not part of the scope are not taken into account.

2.2.2 Exploratory Review

A systematic approach has two potential weaknesses. Firstly, it can lead to the exclusion of several relevant publications and, secondly, some papers refer to synchromodality with (slightly) different names. Therefore, an exploratory approach was used to find relevant papers that have been missed with the systematic approach. The exploratory search approach primarily involved following up references cited by the publications identified during the systematic review and reviewing papers suggested by experts in the field. The criterion for the inclusion of found papers is that the paper contributes additional insights compared to the found papers in the systematic review. The search involved following up references cited by the publications identified, reviewed and categorized during the systematic review.

2.3 Synchromodality

The first subsection will answer the sub-questions:

SQ1a: What is the difference of synchromodality in comparison with other types of transportation?

SQ1b: Why need logistics providers an integration platform?

In 2.3.2 synchromodality is introduced by giving the definition and explaining what synchromodality is. Section 2.3.3 elaborates about integration platform in the logistic sector.

2.3.1 Literature Results

This subsection shortly describes the findings of the literature study. The results of the systematic review are discussed before the exploratory review results are discussed.

To find definitions and the usage of synchromodality in the field, two general search terms are used. The terms are synonyms so all papers with synchromodal subjects are covered. The search term "synchromodality OR synchromodal" yielded 124 results (114 from Google Scholar and 10 from Scopus). After removing duplicates and publications that were not available to read and download 87 papers remained. These papers are filtered by analyzing the abstracts, the key themes and in what context synchromodality is used in the paper resulting in 32 papers that are all available to read in full text. Many papers mention synchromodality but do not give a definition or use it in another context than what is tried to answer in this section.

A challenge faced in this literature review is that synchromodality is a relatively new concept. The papers used that describe synchromodality are all published after 2011 and have a low number of citations. The paper with the most citations using the concept synchromodality is by SteadieSeifi et al. (Steadieseifi et al., 2014) (50 citations). All the other papers/theses/books have between 0 and 5 citations, with an exception of 3 papers that have respectively 8, 10 and 14 citations. It suggests that synchromodality is not yet a mature scientific topic and it is, therefore, hard to predict what literature is best and most useful. During the exploratory review, five additional papers are added. This brings the total of reviewed literature to 37 papers.

The number of papers and citations on synchromodality papers is low. This can mean two things. Either the definition is not used worldwide or the synchromodality concept is not popular among researchers and the logistic sector. Synchromodality is a method within logistic companies to increase their efficiency, quality of work and service to the customer (Egberink, 2015; van der Burgh, 2012). This means that the logistic sector is interested in the concept of synchromodality but that the definition is not commonly used in scientific papers.

Synchromodality includes the real-time aspect of logistics but in scientific papers this is not always called synchromodality. Researchers call synchromodal also multimodal transport with real-time features. To include this research, an extra search query is added to get a complete as possible picture of the current state of the art. The query “multimodal AND "real-time monitoring" AND logistic” results in 587 papers and books at Google Scholar and only a single result in Scopus. Since synchromodality is a new concept in logistics, chosen is to limit the findings of the current query to only recent papers from 2011. This results in 294 papers or books at Google Scholar and one paper in Scopus. After removing the papers that are not available or in the right context only 18 papers remained. For these papers as well the number of citations is very low, with the highest citation rate of 8 (Riessen, Negenborn, Dekker, & Lodewijks, 2013b).

Not all the papers that are considered relevant or within the scope of the thesis are used since these papers referenced to papers in the study and are therefore removed. All the papers that are used in this section are mentioned in table 2.

2.3.2. What is Synchromodality

Synchromodal transportation is the next step after multimodal, intermodal and co-modal transportation. The difference between multimodal, co-modal and synchromodal transportation is shown in figure 2 (European Container Terminals, 2011).

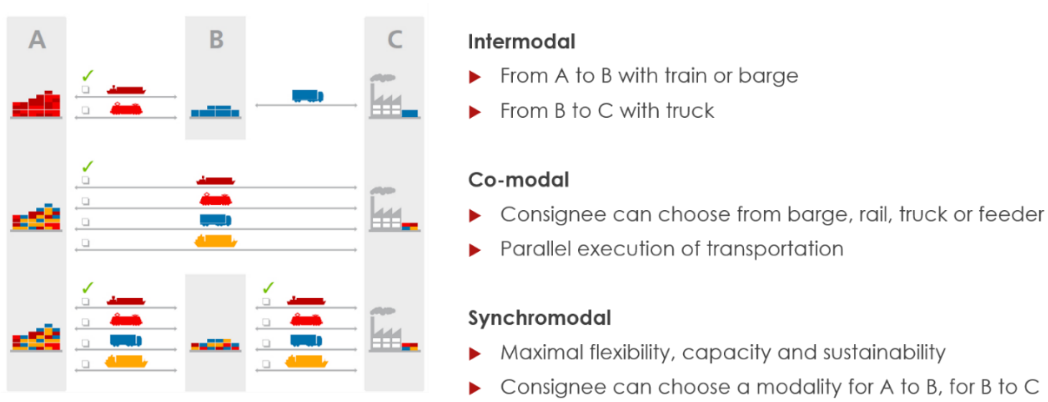

Figure 2: Difference between multimodal, co-modal and synchromodal (European Container Terminals, 2011)

Table 2: Papers used in Section 2.3

Subsection	**Paper**	**Subject**
2.3.2. Synchromodality	Forum, I. T. (2010). *Illustrated Glossary for Transport Statistics*	Definition Multimodal and Intermodal Transportation.
	Steadieseifi et al. (2014). *Multimodal freight transportation planning: A literature review*	Definition Co-modal and Synchromodal Transportation.
	Tan et al. (2011). *The Data Pipeline*.	Definition Synchromodality.
	Riessen et al. (2013). *Analysis of impact and relevance of transit disturbances of planning in intermodal container networks*, 1–22	Definition Synchromodality.
	Baranowski, L. (2013). *Development of a Decision Support Tool for Barge Loading*.	Definition Synchromodality.
	Egberink, J. (2015). *The influence of trust on inter-organizational sharing in logistic outsourcing relationships.*	Benefits Synchromodality.
	van der Burgh, M. (2012). *Synchromodal transport for the horticulture industry*	Characteristics Synchromodality.
	Li et al. (2013). *A general framework for modeling intermodal transport networks*.	Requirements Synchromodality.
	Pleszko, J. (2012). *Multi-variant configurations of supply chains in the context for synchromodal transport*.	Changes by Synchromodality.
	Lucassen, I., & Dogger, T. (2012). *Synchromodality Pilot Study - Identification of bottlenecks and possibilities for a network between Rotterdam, Moerdijk and Tilburg.*	Pilot-study Synchromodality
	European Container Terminals. (2011). *De toekomst van het goederenvervoer*.	Difference Figure
2.3.3. An integration platform for logistics	Toth, M., & Liebler, K. M. (2013). *Real-Time Logistics and Virtual Experiment Fields for Adaptive Supply Networks*	Problems in Transportation sector.
	Vivaldini et al. (2012). *Improving Logistics Services Through the Technology Used in Fleet Management*.	Characteristics for Integration Platform.
	ALICE WG2. (2014). *Corridors , Hubs and Synchromodality Research & Innovation Roadmap*.	Impacts corridors, hubs and Synchromodal
	Cabanillas et al.. (2014). *Towards the Enhancement of Business Process Monitoring for Complex Logistics Chains.*	Challenges
	Zaiat, A., Rossetti, R. J. F., & Coelho, R. J. S. (2014). *Towards an Integrated Multimodal Transportation Dashboard*.	Example dashboard combining all modalities
	Oude Weernink, M. J. P. (2015). *Development of an Integration Platform Metamodel*	Integration Platform Meta-model

Before introducing the concept of synchromodal transportation, the differences between the three types of transportation are explained:

Multimodal Transportation: Multimodal transportation is the transportation of goods by at least two different modalities of transport (Forum, 2010). Mostly a container is transported but it can also be a box or trailer. During the change of modality, all the goods are fully switched to the next modality. The goods will be packed again or bulked.

Intermodal Transportation: Intermodal transportation is defined as a type of multimodal transportation where the load is transported from an origin to a destination in one and the same transportation unit without handling of the goods themselves when changing modes (Forum, 2010). During the change of modalities, the goods remain in the same package. This package can be, for example, a container or a trailer.

Co-modal Transportation: Co-modal transportation focuses on the efficient use of different modes of their own and in combination. It has two differences from multimodal transportation: First, it is used by a group of the shipper in the chain and, secondly, transportation modes are used in a smarter way to maximize the benefits of all modes, in terms of sustainability (Steadieseifi et al., 2014).

A clear overall definition has not been yet adopted. Synchromodal transportation involves a structured, efficient and synchronized combination of two or more transportation modes (Steadieseifi et al., 2014). Through synchromodal transportation, the carriers or customers select independently, at any time, the best mode based on the operational circumstances and/or customer requirements. Tan et al. describe synchromodal transportation as "*a flexible and sustainable transport system in which companies can make an intelligent choice from a range of transport modes modalities*" (Tan et al., 2011). Riessen et al. use synchromodal transportation as "*intermodal planning with the possibility of real-time switching between the modes*" (Riessen, Negenborn, Dekker, & Lodewijks, 2013a). The definition used by Baranowski (Baranowski, 2013) describes it as "*is the optimal, flexible, and sustainable usage of different transport modalities in a network under direction of a logistics service provider, such that the customer is offered an integrated solution for its transportation.*" This is quite similar to the definition of CAPE Groep introduced in chapter 1.

In this thesis, the definition of van der Burgh will be used (van der Burgh, 2012). Synchromodality is described as *"transport where customers agree with a logistics service provider for the delivery of products at a specified cost, quality, and within sustainability targets, but gives the logistics service provider the freedom to decide how to deliver according to those specifications".*

Benefits of synchromodality

Synchromodal initiatives can provide multiple benefits to all the parties involved in the supply chain. This could be the delivery of an improved service through higher frequency by the usage of more effective logistic flows, reduced operational risk, better exchange of knowledge and people between cooperating parties, reduced stocks, reduced CO_2 emission and reduced costs (Egberink, 2015). To describe synchromodality, Van der Burgh identified four characteristics of synchromodal transport (van der Burgh, 2012):

1. <u>A-modal booking:</u> A-modal booking is the practice of placing a transport order in which the shipper only specifies the location and time of delivery, not the modality by which the transport is carried out.
2. <u>Synchronization:</u> Synchronization is the adjustment of production, warehousing and transport to one another to limit waiting time. Synchromodal transport includes synchronization between the infrastructure, transport services, the transport demands of carriers and shippers.
3. <u>Parallel availability of modalities:</u> At least two different modalities are simultaneously available in synchromodal corridors. This allows the flexible utilization of the various modalities and capacity, also when transport is already underway.
4. <u>Bundling:</u> By combining the products of different origins during transport, better use can be made of the available vehicle capacity.

From a customer perspective, synchromodal transport means that a customer agrees with a logistics service provider (LSP) for the delivery of products at a specified cost, quality, and within sustainability targets, but gives the LSP the freedom to decide how to deliver according to those specifications. This freedom gives the LSP the possibility to deploy different modes of transportation flexible. By integrating and coordinating the use of different transport modalities available, synchromodal transport provides a LSP the opportunity to obtain an optimal use of the physical infrastructure (van der Burgh, 2012).

The decision to switch to different modes of transportation may depend on actual circumstances, such as traffic information, instant availability of assets or infrastructure and all other factors that might change requirements. This requires the assistance of information providers by making the latest logistics information (e.g. transport demands, traffic information, weather information, etc.) available (Li, Negenborn, & De Schutter, 2013). Such proactive planning implies that LSPs use more and more push instead of the traditional pull (Pleszko, 2012). Containers no longer remain at the deep-sea terminals in anticipation of action on the part of the recipient (pull) but are directly moved by barge or train to the inland terminals in the hinterland in a pro-active fashion (push).

In 2011, a synchro-modal pilot study was successfully launched on the corridor from Rotterdam to Tilburg (Lucassen & Dogger, 2012). The study included a network with mode-free booking by shippers, where for each container the best transport lane is selected and where different parties work together in optimizing a transport chain. The results, as shown in Figure 3, indicates the modes truck, barge and rail for the Rotterdam port in 2010, the target for 2033 and the results for the network in this pilot study.

	2010	2033 target	Synchro pilot
	57%	35%	19%
	33%	45%	46%
	10%	20%	35%

Figure 3: Results of the pilot study of the modes truck, barge and rail (Lucassen & Dogger, 2012)

2.3.3 An Integration Platform for Logistics

Currently, lack of transparency is a major problem in supply networks (Toth & Liebler, 2013). Simple questions like ''Where is container now?'' or ''How long will it take for my shipment to be available in my warehouse?'' can often not be answered. The main reason is a very heterogeneous IT-infrastructure and insufficient communication between supply chain partners. Furthermore, simple basic information regarding the current status and geographic position of an object in a supply network is often nonexistent or unavailable for partners or logistics service providers (Toth & Liebler, 2013). Such questions can be answered via an integration platform that monitors real-time shipments about their status and whereabouts.

Vivaldini, Pires and Souza identified four variables, besides reducing costs, that give more emphasis to the ways in which a platform tends to be characterized as the possible distinguishing feature in LSP for those logistics services provided (Vivaldini, Pires, & Souza, 2012):

1. Generating greater service reliability.
With information being generated in real time, it is possible to meet different customer requests, such as locating a delivery, show if there have been failures and generate mitigation actions in advance.

2. Making the client more dependent on the LSP.
By investing in fleet technologies creating information relative to delivery and customer service provided, it is assumed that the LSP incorporates an apparatus of solutions that would be difficult for the client to undertake.

3. Favoring the integration with the client.
To provide customer information in real time, or generate reports to inform them of the status of their deliveries and any other occurrences, is a distinguishing feature because this integrates the services provided with that of the client's business

4. Fleet and delivery information management qualifies an LSP.
Having technology does not mean the ability to respond to the needs of or improve customer services. The LSP needs to generate reliable information that complements and streamlines its management, transforming it into a distinguishing feature capable of providing operational results.

The Alliance for Logistics Innovation through Collaboration in Europe (ALICE) has made a roadmap to achieve in the EU a co-modal transport services within a well synchronized, smart and seamless network, supported by corridors and hubs, providing optimal support to supply chains (ALICE WG2, 2014). They summarized the expected impact of research and innovation activities on corridors, hubs and synchromodality. They split the impacts over the People (e.g. higher customer satisfaction), Planet (e.g. lower CO_2 emissions) and Profit (e.g. lower total costs over the supply chain) and split the impacts into primary and secondary impacts. Figure 4 shows the summary of the expected impacts.

An application for monitoring shipments cannot be built without any challenges. Cabanillas et al. identified three important challenges by creating an application that combines all modalities (Cabanillas, Baumgrass, Mendling, Rogetzer, & Bellovoda, 2014).

	Primary Impacts	Secondary impacts
People	+ Customer satisfaction + Products availability + Secure societies	+ Load factors: weight and cube fill of vehicles + Volume flexibility (Time to +/- capacity) + % Synchromodal + Asset utilization + Supply Chain Visibility + Reliability of transport schedules + Perfect order fulfilment + Transport routes optimization (reducing distance) + Transport actors using automatic data exchange + Cargo and logistics units integrated in the automatic data exchange + Upside / Downside Supply Chain Adaptability and Flexibility + Decoupling logistics intensity from GDP - Empty Kilometres - Waiting time in terminals - Risk factor reduction - End-to-end transportation time - Transport distance to reach the market - Lead times
Planet	- Energy consumption (kWh Logistics/GDP) + Renewable energy sources share - CO_2 Emissions (kg CO_2/tKm)	
Profit	+ Return on assets and working capital - Cargo lost to theft or damage - Total supply chain costs	

Figure 4: Summary of expected impacts (ALICE WG2, 2014)

The **first challenge** relates to a gap between how transportation operations can be observed and how state changes are typically represented in business process models. This means that there is a vehicle on the move to transport goods but there is a list of static states. The challenge is here to appropriately align continuous stream of location information with state changes.

The **second challenge** is that logistics operations provide an extensive amount of data which can be hard to relate the data to a specific activity.

The **third challenge** relates to the fact that cargo might be bundled, unbundled and re-bundled during a multimodal transportation activity. The challenging case is to keep track of which container is loaded on which vehicle or vessel.

2.3.3.1. Building an integration platform for logistics

Deciding quickly and reliably are key factors for the successful management of supply networks. This requires real-time information about the current situation and anticipated future behavior in the supply chain. Furthermore, the actors in distribution networks need fast and reliable decision support system (Toth & Liebler, 2013). A logistic integration platform that real-time automatically monitors shipments is a solution and a tool for planners to support their decisions or start certain actions, for example, clear a container at customs the moment it arrived at land. Zaiat et al. propose a solution for the monitoring of the operational state of multimodal transportation systems in a single dashboard by combining all modalities (Zaiat et al., 2014). They state that a decision support system using a dashboard has the potential to be helpful and successful in the logistics sector but some future that has to be done before such platform or dashboard can be successful. They state that a further specification of the data model is necessary depending on practical needs and applications meaning that such a data model is not common yet for all modalities. Furthermore, the implementation of interfaces with external data sources, such as communication protocols, has to be researched and a specification of mapping rules for raw data and database fields has to be designed.

Oude Weernink describes the design of a domain independent meta-model for integration platforms and the application of this meta-model in the logistic domain (Oude Weernink, 2015). The meta-model includes all the seven concepts that are necessary to design a logistic integration platform. A central part of the integration platform is data transformation which transforms data from one source in such a way that it can be handled by the rest of the integration platform. A CDM provides the ability to transform easily from one format to another and, therefore, foster communication between end users and with data sources. Data transformation is one of the seven concepts in the meta-model. The other concepts are:

- **Information interface:** User interfaces for the different users.
- **Service bus:** It has the responsibility to send requests to the right part of the integration platform.
- **Core logic:** Logic in the integration platform that enables the main functions of the platform.
- **Functional supporting modules:** Smaller parts of the platform that support the core logic.
- **Non-functional supporting modules:** Functionalities of the platform that not directly support the core logic. Examples are the administration and authorization modules.
- **Data storage**: The storage of the data in the integration platform.

Figure 5 gives a graphical representation of the meta-model.

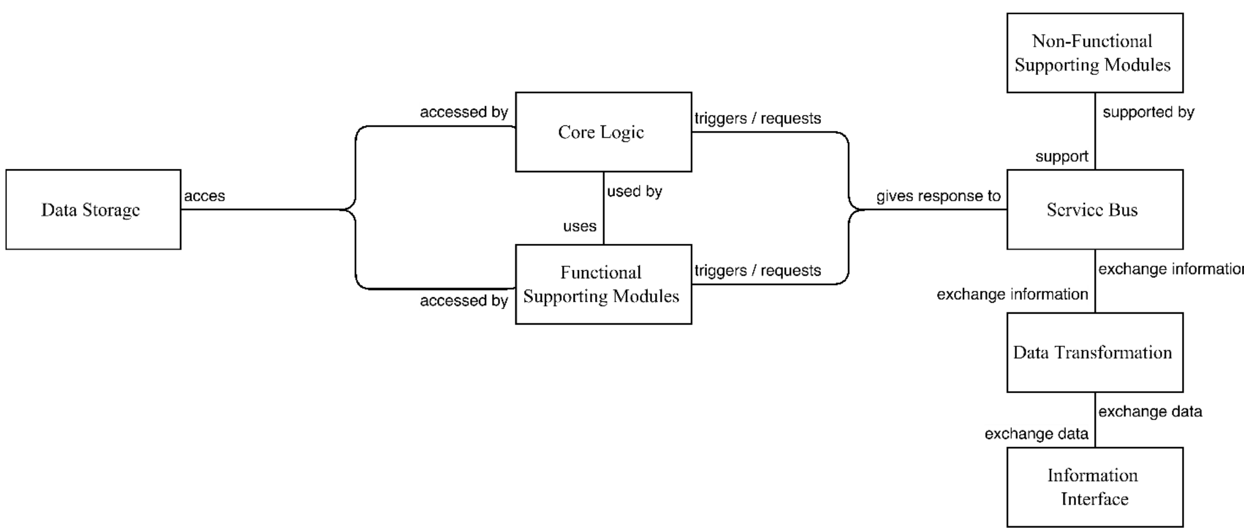

Figure 5: Meta-model for integration platforms (Oude Weernink, 2015)

2.4 The Common Data Model

This section answers the sub question:

SQ1c: How is a common data model for logistics designed?

The first subsection shows the results of the literature search. The following subsections will discuss semantic interoperability, what a CDM is, different methods and what is currently done in current literature in the context of this thesis.

2.4.1 Literature Results

For the search of relevant papers the search query used in Scopus is “"Common Data Model" OR "Canonical Data Model" AND integration”. For Google Scholar, the query “"Common Data Model" "Canonical Data Model" AND integration” is used since the AND in the query does not have the same meaning as in Scopus. In Scholar, every separate word is defined automatically as an OR. In the literature, two definitions for a CDM are used. Common Data model and Canonical Data Model are synonyms (Mangisengi, Huber, Hawel, & Essmayr, 2001) so both words are used in the query to get a complete search. In the context of this thesis, the CDM is used as a central part of an integration bus. Therefore, the goal of the search is to find relevant papers that are related to designing a CDM that is used for integration purposes.

The search terms yielded 256 results (120 from Google Scholar and 136 from Scopus). After removing duplicates and publications, filtered by analyzing the abstracts, the key themes and in what context of this thesis is used in the paper resulting in 17 papers that are all available to read in full text. During the exploratory review, 17 papers are added. This brings the total of reviewed literature to 34 papers.

Not all the papers that are considered relevant are used. All the papers that are used in this section are mentioned in table 3.

Table 3: Papers used in section 2.4

Subsection	**Paper**	**Subject**
2.4.2. Semantic Data Operability	Brandt, P., Basten, T., Stuijk, S., Bui, V., Clercq, P. De, Pires, L. F., & Sinderen, M. Van. (2013). *Semantic interoperability in sensor applications Making sense of sensor data.*	Interoperability
	Bishr, Y. (1998). *Overcoming the semantic and other barriers to GIS interoperability. International Journal of Geographical Information Science.*	Levels of Interoperability
	Böhmer, M., Schmidt, M., & Weißenberg, N. (2015). *Cloud Computing for Logistics. Seamless Interoperability in Logistics: Narrowing the Business-IT Gap by Logistics Business Objects*	Interoperability in Logistics
	Ouksel, A., & Sheth, A. (1999). *Semantic Interoperability in Global Information Systems.*	Ontology Definition
	Longhorn, R. (2006). *Geospatial Standards, Interoperability, Metadata Semantics and Spatial Data Infrastructure.*	Translation Ontology to XML
2.4.3. Model, Data model and Common Data Model	Booch, G. (1998). *The Unified Modeling Language. Techniques.*	Definition model and data model.
	Teorey et al.. (2006). *Database Modeling and Design.*	Definition Data Model.
	West, M. (1999*). Developing High Quality Data Models.*	Definition Data Model.

	Zhang, J. (1994). *A Formal Specification Model and Its Application in Multidatabase Systems*	Definition Common Data Model
	Mangisengi et al. (2001). *A Framework for Supporting Interoperability of Data Warehouse Islands Using XML*	Definition Common Data Model
	Sheth, A. P., & Larson, J. a. (1990). *Federated database systems for managing distributed, heterogeneous, and autonomous databases*	Definition Common Data Model
	Singh, P. M., & van Sinderen, M. (2015). *Interoperability challenges for context-aware logistic services - The case of synchromodal logistics.*	Definition Common Data Model
2.4.4. Design of a CDM	Weske, M. (2007). *Business Process Management*.	Design Data Model
	Chen, P. P.-S. (1976). *The entity-relationship model---toward a unified view of data.*	E-R approach
	Booch, G. (1998). *The Unified Modeling Language. Techniques.*	UML Approach
	Lampathaki et al. (2009). *Business to business interoperability: A current review of XML data integration standards.*	Data modeling concept
	Böhmer, M., Schmidt, M., & Weißenberg, N. (2015). *Cloud Computing for Logistics. Seamless Interoperability in Logistics: Narrowing the Business-IT Gap by Logistics Business Objects*	Method. Data Modelling
2.4.4.1. Data Models and Architectures	TOGAF. (2009). *The open group architecture framework, version 9, enterprise edition.*	TOGAF
	Iacob, M.-E. et al. (2012). *Delivering Enterprise Architecture with TOGAF and Archimate.*	Archimate
	Iacob, M. (2013). *Sct architectuur*	Method Data Modelling.
	Iacob, M. E. et al. (2013). *Towards a reference architecture for fuel-based carbon management systems in the logistics industry*	Method Data Modelling and Architecture
	Niemi, E. (2006). *Enterprise Architecture Benefits : Perceptions from Literature and Practice*	Enterprise Architecture
2.4.4.2. Standards and Data Models	Sliwczynski et al. (2012). *Standards for Transport Data Exchange in the Supply Chain – Pilot Studies*	Overview Standards in Logistics
	Pedersen, J. T. et al. (2011). *One common framework for information and communication systems in transport and logistics.*	Common Framework
	Golinska, P., & Hajdul, M. (2012). *Virtual Logistics Clusters – IT Support for Integration Introduction - Virtual Logistics Clusters.*	Example usage Common Framework
	GS1. (2007). *Logistics Interoperability Model*	Example CDM
	Iacob, M. E. et al. (2013). *Towards a reference architecture for fuel-based carbon management systems in the logistics industry.*	Example CDM
	Iacob, M. (2013). *Sct architectuur*	Example CDM
2.4.4.3. Conflicts	Sheth, A. P., & Larson, J. a. (1990). *Federated database systems for managing distributed, heterogeneous, and autonomous databases*	Explanations Conflicts
	Zhang, J. (1994). *A Formal Specification Model and Its Application in Multidatabase Systems*	Categorization Conflicts
2.4.5. Transformation and Mapping	Bray et al. (2008). *Extensible Markup Language (XML)*	XML
	Liam, Q. (2015b). *XML Schema - W3C*	XML Schema
	Chung, S. M., & Mah, P. S. (1995). *Schema integration for multidatabases using the unified relational and object-oriented model.*	Importance Semantic Heterogeneity

	Liam, Q. (2015a). *Transformation - W3C*	Transformation Definition
	Domínguez, E., Lloret, J., Rubio, Á. L., & Zapata, M. A. (2005). Evolving XML schemas and documents using UML class diagrams	Mapping

2.4.2. Semantic Data Operability

In logistics, processes and supply chains are highly dynamic and the business requirements are always changing (Hendrik Haan, van Hillegersberg, de Jong, & Sikkel, 2013). Processes and supply chains involve a number of parties and rely on a variety of IT systems and applications of different providers. These constantly changing requirements create challenges to establish and maintain the processes meeting the business requirements on both the business level and technical level.

In the supply chain, applications deliver data that can have distinct meanings in different domains. Brandt et al. argue that semantic interoperability requires not only sharing of data and concepts but also positioning them in the specific purpose of the application for their appropriate use (Brandt et al., 2013).

In order to create a platform that is connected with all kinds of applications, and create an understanding of the data and concepts between different applications interoperability is required. Ontologies have been acknowledged in the literature as a means to achieve interoperability between systems at the semantic level (Ouksel & Sheth, 1999). Brandt et al. define interoperability as the ability of an application or product to work with other application or products in a loosely coupled network on the part of the customer (Brandt et al., 2013). Semantic interoperability relates to sharing a conceptualization by communicating its representation as a message that, with minimal mutual dependency, leads to a mutual appropriate interpretation. This means that when applications communicate in a loosely coupled platform they have the same understanding and interpretation of the contents of the messages.

Bishr defined six levels of interoperability between two or more systems or applications (Bishr, 1998). The six levels are shown in Figure 6:

The first level is network protocols which let users communicate without any direct service from the applications to merely download flat files.
The second level provides a higher level network protocol where users can connect to a host and interact seamlessly regardless of its operating system.
The third level of interoperability is the exchange of spatial data files. Users can download data files in a standard format and the system automatically identifies, and consequently converts, the format to that of the user.
The fourth level allows users to establish a connection between their own application and others and once these are established they can query the remote applications using their own, local, query language.
The fifth level provides users with a single global data model that is an abstraction of all the underlying remote databases. Queries are sent to the global model that in turn maps between the user's and the remote data models. Although users can query remote databases seamlessly, knowledge on their semantics is still required to reach interoperability. As mentioned before, two applications in different domains can have the same data model, but they vary in their assumption and semantics.
The sixth level provides seamless communication between applications without having prior knowledge of their underlying semantics.

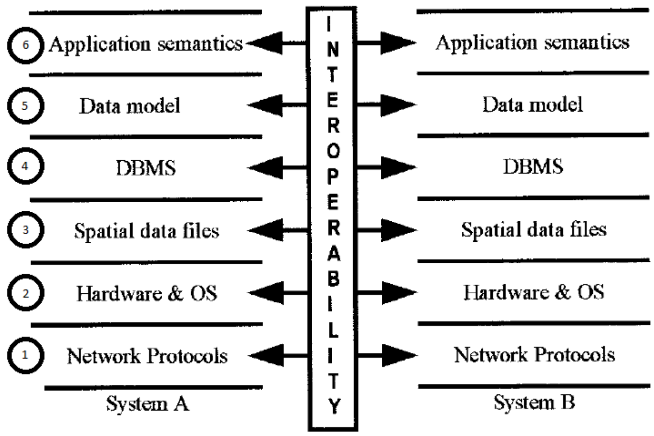

Figure 6: Levels of interoperability (Bishr, 1998)

Nowadays, the rapid pace that business requirements change in logistics requires flexibility and modularity of (physical) logistics services and IT systems. Hence, extensive IT systems are more and more broken down into smaller software applications (Böhmer, Schmidt, & Weißenberg, 2015). Significant competitive advantages are reached when these applications are seamless interoperable. Standardization is necessary on both levels in order to narrow the gap and accelerate process adaptations for communication between companies and between the business and IT people within the company.

To close the gap between business and IT a common canonical (data) model is required. However, since IT and Business have a different focus, different views of the common canonical (data) model should be supported (Böhmer et al., 2015). To close the communication and interpretation gap between companies Böhmer et al. introduce an independent application that uses the common data model in a loosely coupled platform. This independent application communicates with other parties in the platform according to the common data model. Messages are translated back to the other systems so that these systems can interpret these messages. A bus system is necessary for translating and directing the messages. This is shown in Figure 7.

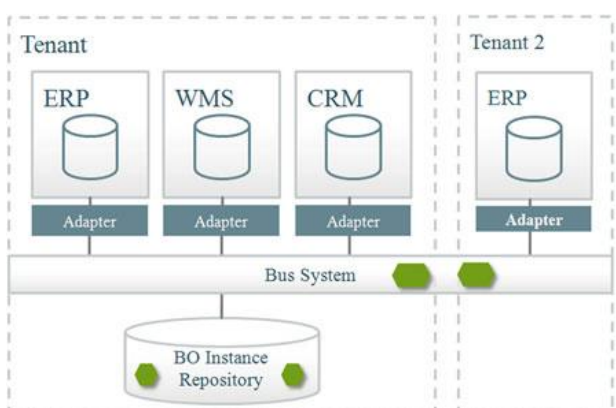

Figure 7: Scenario mentioned by Böhmer et al. (Böhmer et al., 2015)

2.4.3. Model, Data Model, Common Data Model

A model is a representation in a certain medium of something in the same or another medium. A 3D model or a drawn overview of a building on a sheet of paper are both examples of a model. The model captures the important aspects of the thing modeled from a certain point of view and simplifies or omits the rest (Booch, 1998). It is used to capture and precisely state requirements and domain knowledge so that all stakeholders may understand, discuss and agree on them.

In Software Engineering, a data model is used as a conceptual schema, meta-model, that shows all the data, their relationships and in what form data are presented and are used as a starting point for interface or database design (Teorey, Lightstone, & Nadeau, 2006). It is intended to be easier to use than the final system (Booch, 1998). It should be both understandable for business people as for IT people. Data models are developed based on the data requirements or/and the context of an activity model for the system or service that is being developed (West, 1999). A data model will normally consist of entity types, attributes, relationships, integrity rules, and the definitions of those objects.

The concept of a common data model (CDM), also called canonical data model, is central to the process of integration. Mangisengi et al. describe the common data model as a model, which is used for expressing all schema objects available in the different local schemas (of other systems) (Mangisengi et al., 2001). In order to integrate different systems with each their own local data model, the local data schemas have to be transformed into common schemas that are expressed in a CDM. The CDM is used to represent the shareable knowledge of each component database system and, as such, must have at least the same representational power as any of the local data models (Zhang, 1994). It expresses all the schema objects available in the different local schemas (Mangisengi et al., 2001). Sheth and Larson considered that the CDM should be semantically rich; that is, it should provide abstraction and constraints so the to the semantics relevant to schema integration can be presented (Sheth & Larson, 1990).

In integration projects, a CDM brings various data models together. It is a standardized definition of how system solutions and technologies represent resources and their relationships. It solves the problem that data are represented with different encodings, use various data formats, and are exchanged with different protocols (Singh, van Sinderen, 2015). The CDM acts as the standard for consistently describing details and identity of resources and is used as the basis of data modeling for interactive/integrated designs. Because the CDM is an information model, solutions are able to maintain existing database schemas. When integrating, CDM definitions and terminology are in use. This fosters consistent that is re-usable across multiple solutions.

2.4.4. Designing a Common Data Model

The design of a common data model is the basis for the data integration of heterogeneous data (Weske, 2007). This means that data with the same structure from different sources could be integrated into the same database. Different notations and languages to design a common data model exist. Examples are the Entity Relationship (Chen, 1976) and the Unified Modelling (Booch, 1998).

One of the languages to classify and organize data in a given domain is the Entity-Relationship approach (Chen, 1976). Entity-Relationship approaches are used in structured analysis and conceptual modeling. They are often used to graphically represent the logical structure of a database. Another language is the

Unified Modelling Language (UML) (Booch, 1998). UML is a standardized modeling language that is primarily used for object-oriented software. It consists of a family of graphical notations that assists in describing and designing software systems. The UML is independent of an implementation language. In summary, UML is a popular and standardized modeling language that is primarily used for object-oriented software. Entity-Relationship diagrams are used in structured analysis and conceptual modeling. They are often used to graphically represent the logical structure of a database.

Another way of data modeling is the concept of Lampathaki et al. as shown in Figure 8 (Lampathaki, Mouzakitis, Gionis, Charalabidis, & Askounis, 2009). They state that data models exist at multiple levels:

• **The Conceptual Data Model** describes data from an abstract yet semantically-enriched point of view. It includes ontologies, classes and their attributes.

• **The Logical Data Model** provides more details than the conceptual data model and creates the data structures that will be exchanged during the transactions. It includes core components, the individual pieces of information, and the XML Schemas with which the XML data files transferred comply.

• **The Physical Data Model** describes the storage of data in a physical relational database. It is the blueprint for database schemas

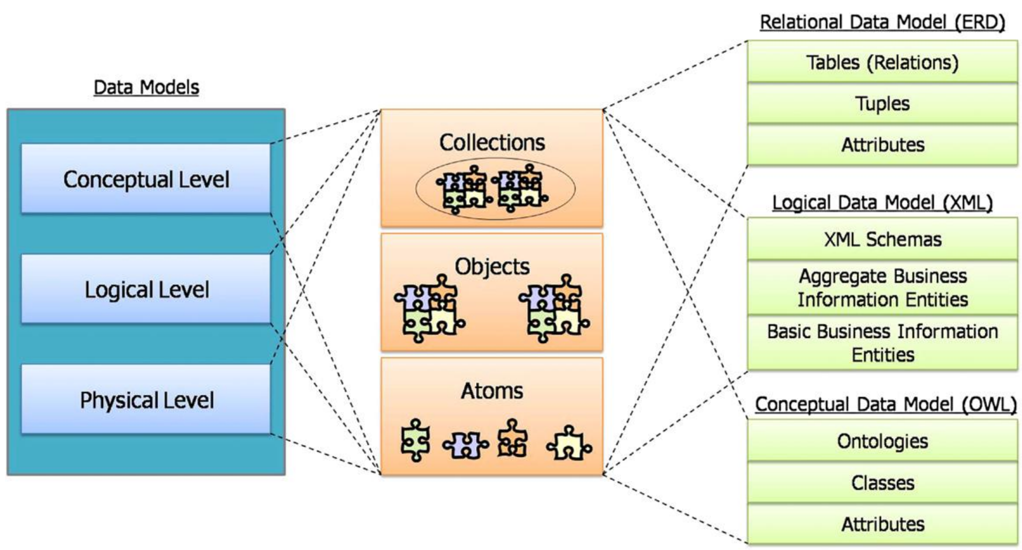

Figure 8: Data modelling granularity levels (Lampathaki et al., 2009)

Lampathaki et al. state that data modeling consist of three steps. First, the designer has to ask <u>why</u> things work and to look at the concepts and the semantics of the data. Second, the designer has to progress to a more logical model by recognizing <u>what</u> data needs to be transferred and its structure. Third, the designer creates the physical models that reflect <u>how</u> data are stored.

To build a data model containing all the business object in the selected domain and identifying their connections Böhmer et al. (Böhmer et al., 2015) developed a domain model development methodology. This methodology is a four phase based and iterative approach to built a model that specifies all real and virtual entities that are relevant in a selected domain and structures them by sets of attributes, reusable sub-components and relationships to other objects.

Step 1: Problem and Requirements Definition – In this stage the problems in the selected domain are identified and requirements for the model are designed.

Step 2: Analysis – The goal of this step is to identify all the objects within the selected domain. The preferred method for the identification is the usage of scenarios. Based on these scenarios the objects, their attributes and relations can be identified.

Step 3: Design – The third step is to model the objects and its associations and relations and the definition of all the attributes related to the object. This results in a first domain model.

Step 4: Maintenance – The last step consist of iterative improve the model. It is clear that no model is faultless right from the start. The model can be updated in various ways from feedback from companies to additional object and attributes that were missed or found unimportant during previous sessions.

2.4.4.1. Data Models and Architectures

An integration platform fosters communication and increases the data shared over the supply chain. Such platform provides greater flexibility for accessing and sharing information and resources. The implementation of an integration platform changes not only the enterprise architecture of the logistics providers but for the whole supply chain as well which makes it necessary to describe the relationships and communication in the supply chain. Enterprise architecture includes all the models needed in managing and developing an organization and takes a holistic view of its business processes, information systems and technological infrastructure (Niemi, 2006). Niemi state that it is claimed that enterprise architecture provides a vehicle for aligning and integrating strategy, people, business and technology, and enabling an agile enterprise, continually evolving within the ever-changing environment. A language to model relationships between enterprise architecture and business performance is Archimate.

Archimate is a language that can be used for the development, maintenance and operationalization of processes, organization structures, information flows and IT systems. This is presented in three different layers (M.-E. Iacob, Jonkers, Quartel, Franken, & van den Berg, 2012). The business layer establishes the value proposition, business strategy and the working model of the enterprise. In order to deliver added-value to its environment, the business layer makes use of services delivered by the application layer. Applications support all the process taking place within the business layer. The application components make use of service provided by the technology layer. Technology layer forms the technical infrastructure of the organization.

Iacob, van Sinderen, Steenwijk and Verkroost presented a reference enterprise architecture and data model for a carbon management system (M. E. Iacob, Van Sinderen, Steenwijk, & Verkroost, 2013). The paper shows a step-by-step method for developing an architecture and data model in the logistics domain. The method starts by developing the reference architecture specification using the development methodology prescribed by The Open Groep Architecture Framework (TOGAF, 2009) and using the enterprise architecture modeling language Archimate (M.-E. Iacob et al., 2012). The first step is to define the business process domain. It describes the processes are done at the business level in order to complete the whole business process. The second step is to define the Information Systems domain. It gives an overview of the typical software applications landscape. The third step is to integrate the business and IT domains. If the reference architecture is made and the scope of the application/prototype is defined, the data model can be created. The same approach is used during the design of a Synchromodal Control Tower (M. Iacob, 2013).

2.4.4.2. Standards and Data Models in Logistics

Over the years, there have been multiple approaches which aim for a common standard for communication between collaborating partners in multimodal transport (Sliwczynski, Hajdul, & Golinska, 2012). There are so many different standards that it is not possible to discuss and mention each approach. This subsection discusses a few important research works in multimodal transport.

The first work introduces a common framework as a result of eight different logistic projects in the EU (Pedersen, Paganelli, Knoors, Meyer-Larsen, & Davidsson, 2011). They proposed eight standard messages for any shipment to supports interoperability between commercial actors and communication to authorities and transportation network responsible. These messages are suitable for communication between actors in a synchromodal context since they take into account real-time information. For each standard message a small description is given:

1. Transport Service Description – This is a standard message telling the customer what service can be provided by a LSP.
2. Transport Execution Plan – This message describes all the information needed related to the execution of a transport service.
3. Goods Item Itinerary – It provides information about the schedule of delivery of the goods. This message can be used to check the current location of the goods.
4. Transport Execution Status – The message provides information about the progress of the transport and of the cargo condition.
5. Transport Operation Status – Assists in establishing the best possible arrival time estimates.
6. Security Data Message – Provides information about the security of a sealed load unit.
7. Common Regulatory Schema – Provides a unified way of informing authorities about transport such that compliance may be verified.
8. Transportation Network Status – This message provides the current available capacity of the transport infrastructure.

Golinska and Hajdul designed a comprehensive tool supporting the consolidation of transport processes for a group of independent companies using this common framework to improve cooperation between organizations (Golinska & Hajdul, 2012). Using the tool in a field study they concluded that the tool has the great potential for reducing costs.

GS1 is a worldwide organization that designs and implements standards for the exchange of electronic messages between organizations. They developed a logistic interoperability model where retailers, manufacturers, material suppliers and logistics service providers are represented (GS1, 2007). They specified the messages exchanged by different parties in different phases of the order. For this thesis, the transport phase is most interesting (see figure 9). This is divided into three larger group of messages:

Collect, Monitoring and Deliver. These three groups represent the three larger events in the delivery of goods: The pick-up of the goods, the transportation and the delivery.

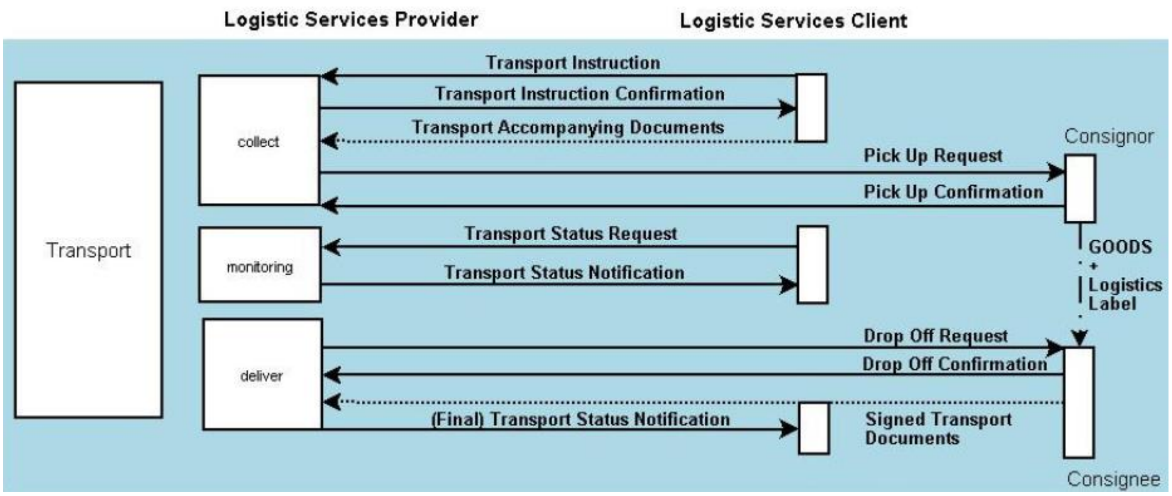

Figure 9: Transportation messages in the logistic interoperability model (GS1, 2007)

The research of Iacob, van Sinderen, Steenwijk and Verkroost presented data model that is focused on a prototype that calculates the fuel-based carbon emission of a trip (M. E. Iacob et al., 2013). They propose an architecture for a system that monitors and collects real-time data about the fuel consumption during trips and calculates detailed and accurate carbon footprints of transportation services. They validate the proposed architecture by implementing it by designing a prototype. However the intention of the research does not fit within the scope of this thesis, the data model is relevant. It considers multimodal transport and frequent data exchanges between collaborating logistic partners as the cornerstone for an efficient logistic solution. The prototype required a data model and includes different entities as the shipment, the shipper, the trip and location which can be used as a reference designing a data model for shipment tracking.

Iacob presented an architecture and a data model for a synchromodal control tower (M. Iacob, 2013). The goal of the document is to design a high-level architecture for a synchromodal control tower using Archimate. The control tower is defined in a data model for the order and the network where particular the definition of an order is interesting.

2.4.4.3. Conflicts

The integration of different data sources and company systems involves multiple databases dispersed over different locations. Each system has their own pre-existing database systems that are autonomous and usually heterogeneous (Sheth & Larson, 1990). Each system is designed and built by different people for different purposes. They typically use different conceptual schemas expressed by the same data model or different data models to describe the real-world object. According to Zhang, these conceptual schemas have a number of discrepancies that can be classified as a schematic, semantic or data conflict (Zhang, 1994).

A **schematic conflict** occurs because of the different types of how a data model can be presented. For example, a relation in a relational data model cannot be nested but in object-oriented schemas, an object can contain another object as its attribute.

A **semantic conflict** is caused because different conceptual schemas can have different semantic meanings for the same real-world objects.

A **data conflict** occurs because data are represented in different scales, precisions, units or the wrong data for integration. For example, the age of a person can be represented by its age of by its birth date.

These conflicts must be removed by performing transformations on the schemas to produce equivalent schemas (Poulovassilis & McBrien, 1998).

2.4.5. Transformations and Mapping

In order to send messages between different systems with different input and output messages and data models, transformations have to be made to transform a message into a message that is acceptable by the receiving system. Over the World Wide Web, eXtensible markup language (XML) has emerged as the standard for data exchange (Bray, Paoli, Sperberg-McQueen, Maler, & Yergeau, 2008). Each time an XML document is processed, a check against a specified XML schema can be performed to ensure its validity. The system's schema can be represented in a XML Schema Definition (XSD) file (Liam, 2015b). A XSD file expresses constraints about XML documents that a system can process. For the integration of different system's schemas, a crucial point is to overcome the semantic heterogeneity of the schemas to be integrated (Chung & Mah, 1995).

To transform one XML file into another, a transformation is needed. This transformation is defined in the XSL Transformations (XSLT) language. XSLT is a language for transforming XML documents into other XML documents, text documents or HTML documents (Liam, 2015a). An important aspect of the transformation is mapping. Mapping translates the conceptual components to the logical components (Domínguez, Lloret, Rubio, & Zapata, 2005). An example is shown in Figure 10. The XSLT document describes the transformation of the XML files by specifying which components of one XML file correspondents to the component of the other XML file.

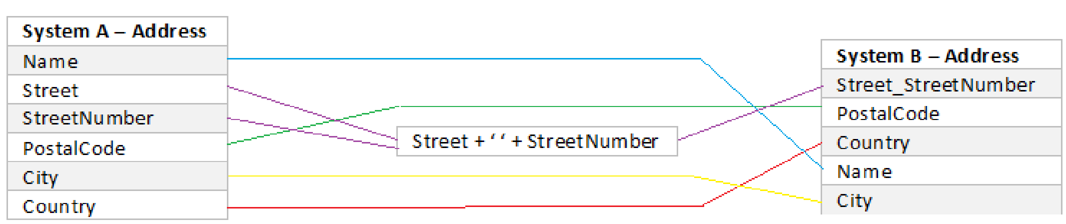

Figure 10: An example mapping of two definitions

2.5 Big and Open Data

The World Wide Web is full data, either structured, semi-structured or unstructured. This section elaborates on these forms, what technique can be used to capture and retrieve this data. It answers the sub-question:

SQ1d: How can big and open data dealt with on the World Wide Web?

The first subsection gives the results of the literature search. The following subsections will discuss what kind of data can be found on the web and what technique and tools can be used to retrieve it.

2.5.1 Literature Results

For the search of relevant papers the search query used in Scopus is "web scraping" AND "data mining" AND "unstructured data"". For Google Scholar, the query “"web scraping" "data mining" "unstructured data"" is used since the AND in the query does not have the same meaning as in Scopus.

The search terms yielded a total of 94 results (94 from Google Scholar and 0 from Scopus). No papers are found in Scopus, but extending the search query results in Google Scholar in an increase of hundreds of papers while the search terms used here fit well in the scope. After removing duplicates and publications, filtered by analyzing the abstracts, the key themes and in what context of this thesis is used in the paper resulting in 5 papers that are all available to read in full text. During the exploratory review 2, additional papers are added. This brings the total of reviewed literature to 7 papers.

Not all the papers that are considered relevant are used. All the papers that are used in this section are mentioned in table 4.

Table 4: Papers used in section 2.5.

Subsection	**Paper**	**Subject**
2.5.2 Data	Blumberg, R., & Atre, S. (2003). *The Problem with Unstructured Data.*	Definition Unstructured Data
	Abiteboul, S., Buneman, P., & Suciu, D. (1999). *Data on the Web: From Relations to Semistructured Data and XML.*	Definition Semi-Structured and Structured Data
2.5.2 Structuring Data	Qureshi, P. A. R., Memon, N., Wiil, U. K., Karampelas, P., & Sancheze, J. I. N. (2011). *Harvesting Information from Heterogeneous Sources*	Definition Information Harvesting
	Baumgartner, R., Gatterbauer, W., & Gottlob, G. (2009*). Web data extraction system.*	Outcomes Information Harvesting
	Vasani, K. (2014). *Content Evocation Using Web Scraping and Semantic Illustration.*	Definition Web Scraping
2.5.4. Cloudscrape	Kanaoka, K., Fujii, Y., & Toyama, M. (2014). *Ducky : A Data Extraction System for Various Structured Web Documents*	Web Scraping Tool
	Veldhuis, R. (2015). *Developing an automated solution for ETA definition concerning long distance shipping.*	Choice tool

2.5.2 Structured, semi-structured and unstructured data

Data stored on the internet is presented in one of the three different kinds: Structured, unstructured and semi-structured. This is shown in figure 11.

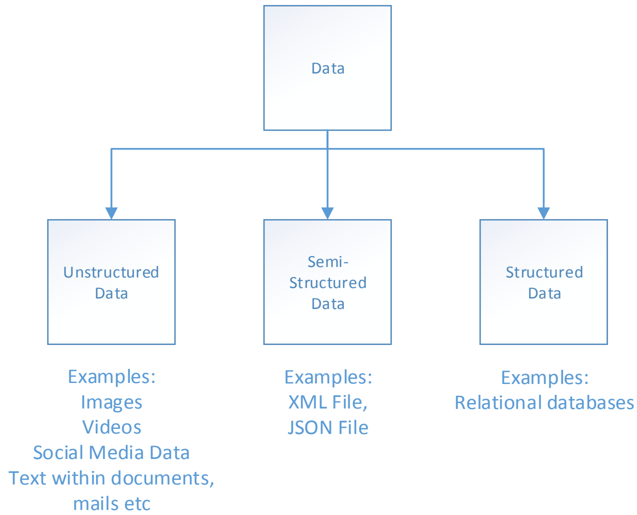

Figure 11: Unstructured, semi-structured and structured data

Unstructured data are data that has no identifiable structure within itself available (Blumberg & Atre, 2003). Unstructured data are also described as data that cannot be stored in rows and columns in a relational database. An example of unstructured data is a document that is archived in a file folder, images and videos. The advantage of unstructured data is, that no additional effort on its classification is necessary. A limitation of this kind of data is, that no controlled navigation within unstructured content is possible.

Semi-structured data are often explained as ”...*schemaless or self-describing, terms that indicate that there is no separate description of the type or structure of the data*” (Abiteboul, Buneman, & Suciu, 1999). Semi-structured data do not require a schema definition. A schema definition is optional, so it is possible to structure data. Data instances do also exist in the case that the schema changes. Furthermore, a schema can be defined according to already existing instances instead of the requirement that the schema is defined before any instances can exist. The types of semi-structured data instances may be defined as a part of the data and it is also possible that a data instance has more than one type. Semi-structured data have the ability to accommodate variations in structure (Abiteboul et al., 1999). This means that data may be created according to a specification or close to a type. An example of semi-structured data is XML. In XML data can be encoded directly and a XML Schema defines the structure of the XML document.

Structured data are structured according to a predefined schema. Abiteboel et al. describe structured data as that conforms to a specification of a schema of that particular data (Abiteboul et al., 1999). A typical example of fully structured data is a relational database system. The design of a schema has to be defined before the data is created and the database is filled. The schema defines the type and structure of data and its relations. A well-defined schema enables efficient data processing and an improved storage and navigation of the data. However a designed database has high performance and navigation, the disadvantages are that it is not very flexible and scalable. For example, it is not possible to extend a single table row with a new attribute without creating another table column.

2.5.3 Structuring data

For a logistic integration platform, data are retrieved from at least two types of sources: Back-office systems of companies and the internet. From the back-office systems, data are received via a semi-structured XML format. The data stored in the back-office system is stored in a structured way but send in a XML format. This XML format is specified of how such systems communicates and can receive and process incoming XML messages.

However the information dispersed over the internet is available in heterogeneous formats such as PDF, HTML and XML. Information noise, advertisement links, and text used for site navigation complicates the task of information harvesting. Qureshi et al specify information harvesting as the "*extraction of useful noise-free information in some context stemming from heterogeneous information sources"* (Qureshi, Memon, Wiil, Karampelas, & Sancheze, 2011). The process of extracting data on the internet is referred as web scraping or web data mining. The information noise creates the problem that the data are mixed together with advertisements or web page specific information such as navigation links. This way the web page is made user-friendly, but for machines, the actual content is hard to retrieve. With information harvesting the output is structured data, for example in a relational database, that enables to analyze and compare data (Baumgartner, Gatterbauer, & Gottlob, 2009). Data should be presented in a simple format that is easy to process and analyze. This is probably the plain text itself without all information added (Qureshi et al., 2011).

Web scrapers simulate the human exploration of the World Wide Web by either implementing low-level hypertext transfer protocol or embedding suitable Web browsers (Vasani, 2014). Web scraping is closely related to Web indexing. Web indexing is a technique for information retrieval adopted by search engines to automatically index information at the internet through a bot. However, web scraping focuses on the transformation of unstructured data on the Web into structured data that can be stored and analyzed in a relational database.

2.5.4 Capturing big and open data

On the internet, various tools are available for web scraping, ranging from scripting languages to GUI guided solutions. According to Kanaoka et al. a CSS selectors is preferred above other web scraping methods (Kanaoka, Fujii, & Toyama, 2014). Another common method is XPath. CSS selection and XPath have many similarities. They both using HTML tags in a generated web page as selection methods but CSS selection fits better with the complexity of modern websites and allows for powerful and robust queries.

Veldhuis conducted a comparison between different web scraping tools (Cloudscrape, RapidMiner, WebHarvest, Mozenda and more) and concluded that CloudScrape (from April 2016 dexi.io) is easy to use, browser-based application with an easy to understand user interface (Veldhuis, 2015). The other tools did either not have an API, was extremely expensive or lacked important features as support for logging in on websites. The web scraping tool used in this thesis is CloudScrape. It support logins on websites, has an API to ease communication with other applications, use the CSS Selection method and is not extremely expensive. Veldhuis used the tool in a previous research conducted at CAPE Groep and meets the requirements that Veldhuis defined (Veldhuis, 2015). In figure 12 the interface of CloudScrape is shown.

The tool follows the actions you go through on a website, saving them and providing the ability to automate them and save certain data from the website along the way.

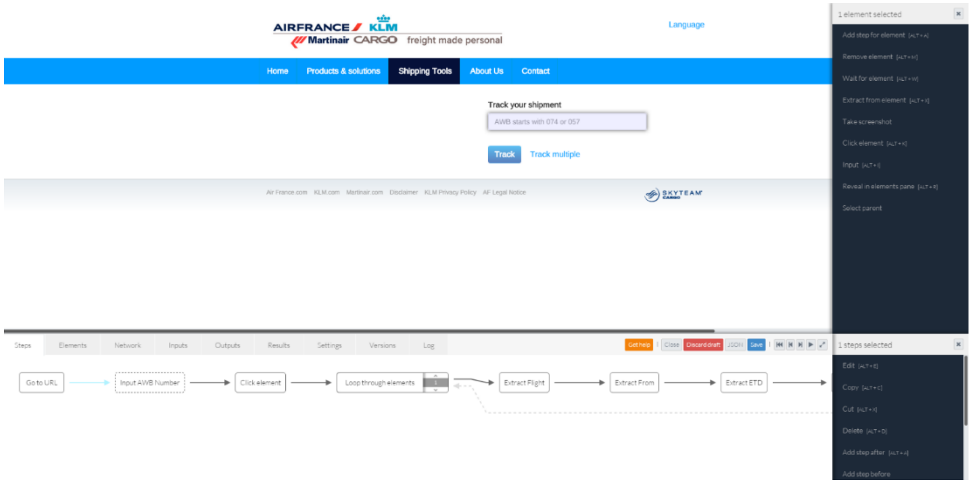

Figure 12: Screenshot of the interface of CloudScrape

2.6 Conclusion

This chapter answered sub-questions SQ1a-d based on the literature studied. Each sub-question will be answered.

SQ1a: What is the difference of synchromodality in comparison with other types of transportation?

In comparison with other types of transportation, synchromodal transportation gives logistics service providers the possibility to select the carriers or customers independently and, at any time, the best modality based on the operational circumstances and/or customer requirements (Steadieseifi et al., 2014). In this thesis, synchromodality is described as *"transport where customers agree with a logistics service provider for the delivery of products at a specified cost, quality, and within sustainability targets, but gives the logistics service provider the freedom to decide how to deliver according to those specifications"* (van der Burgh, 2012). Riessen et al. state that synchromodality includes the possibility of real-time switching between modalities (Riessen et al., 2013b).

To be able to real-time switch, planners require data to make the decision what modality is most appropriate. In the problem statement, in chapter 1, is established that this is currently manually done which is time-consuming. The second sub-question answers why logistics providers have a need to centrally manage this data.

SQ1b: Why do logistics providers need an integration platform?

Currently, simple questions like ''Where is my container now?'' or ''How long will it take for my shipment to be available in my warehouse?'' can often not be answered (Toth & Liebler, 2013). Real-time information about the current situation and anticipated future behavior in the supply chain is necessary in order to decide quickly and reliably. Zaiat et al. did research on monitoring of the status of multimodal transport in a single dashboard by combining all modalities and concluded that a platform with a dashboard showing real-time data has potential to be helpful and successful in the logistics sector (Zaiat et al., 2014). It confirms that logistics have a need to access real-time information via a platform. Before such platform can be designed Cabanillas et al. state that several challenges, for example, the translation from transportation operations into business process models and the bundling and unbundling of cargo during transportation, have to be overcome before shipments can be monitored (Cabanillas et al., 2014). Current processes, decision choices and the impact of a platform on these processes and the organization have to be understood before a platform can be designed.

To create a platform for the logistic sector that enables the usage of real-time data of shipments, Oude Weernink describes the design of a domain independent meta-model for integration platforms and the application of this meta-model in the logistic domain (Oude Weernink, 2015). Oude Weernink states that in an integration platform a common data model is required to foster communication between end-systems.

SQ1c: How is a common data model for logistics designed?

Central in the integration platform is a common data model that fosters communication between systems connected to the platform. A common data model is described as a model, which is used for expressing all schema objects available in the different local schemas (of other systems) (Mangisengi et al., 2001) and is a standardized definition of how system solutions and technologies represent resources and their relationships.

To design a common data model, the method of Böhmer will be used (Böhmer et al., 2015). The method of Böhmer et al. contains four steps:

> **Step 1: Problem and Requirements Definition –** In this stage the problems in the selected domain are identified and requirements for the model are designed.
>
> **Step 2: Analysis** – The goal of this step is to identify all the objects within the selected domain. The preferred method for the identification is the usage of scenarios. Based on these scenarios the objects, their attributes and relations can be identified.
>
> **Step 3: Design** – The third step is to model the objects and its associations and relations and the definition of all the attributes related to the object. This results in a first domain model.
>
> **Step 4: Maintenance** – The last step updates the model in various ways from feedback from companies to additional object and attributes that were missed or found unimportant.

In the common data model as well as to the user the data must be stored and presented unambiguously but data on the World Wide Web is either unstructured, semi-structured or structured and presented in different notations.

SQ1d: How can big and open data dealt with on the World Wide Web?

Böhmer state that data from various sources should be combined and integrated to enable shipment tracking (Böhmer et al., 2015). A challenge is to collect either form of data, transform it to a common format and structure it so users can use this data. During the design of a common data model, the designer has to solve conflicts. A difference in semantics, schema or data can lead to conflicts that must be solved in order to use the data (Zhang, 1994). Transformations in the bus integration system, using a common data model, can solve these conflicts (Poulovassilis & McBrien, 1998). Transformations of messages into the end system or common data model format take care of the harmonization of the data from the World Wide Web. During transformation, the mapping of information translates the conceptual components to the logical components (Domínguez et al., 2005).

Data from the World Wide Web can be structured by information harvesting (Qureshi et al., 2011). With information harvesting the output is structured data, for example in a relational database, that enables to analyze and compare data (Baumgartner et al., 2009). Data should be presented in a simple format that is easy to process and analyze (Qureshi et al., 2011).

The process of extracting data on the internet is referred as web scraping or web data mining. An accepted tool is Cloudscrape (Veldhuis, 2015). Veldhuis concluded that CloudScrape is easy to use, browser-based application with an easy to understand user interface. Other tools did either not have an API, was extremely expensive or lacked important features as support for logging in on websites.

Chapter 3 Interviews, Data and Processes

To identify data sources, data types, processes and needs, observations and interviews were held at Neele-Vat, CTT and HST Groep. This chapter will answer the sub-questions 2 till 4. The findings of these interviews and observations are elaborated in this chapter and is the first chapter in the Building, Intervention and Evaluation step. The chapter starts with an introduction of all terms used in this chapter. Section 3.2 elaborates about the set-up of the interviews. Section 3.3 shows the findings of the interviews and observations. During section 3.4 some example websites are given as well difficulties in data on different websites are discussed. Section 3.5 concludes this chapter.

3.1. Introduction

During this thesis, several companies are visited for interviews with employees and potential users about the current processes and needs and to observe current processes. During these interviews and observations, the track and tracing processes and what information they use and need to make decisions will be identified. The found information creates a basis for the design of the CDM and the processes create a basis for the desired behavior of the integration platform. This chapter shows the results of these interviews and observations. Throughout the thesis definitions are used that are common within the field of logistics but can be confusing for those who are not known to these terms. Table 5 specifies different definitions with an explanation in order to make the definition clearer.

Table 5: Definitions used in section 3 and up

Term	Definition
Import	Goods that are transported to the Netherlands, most of the times the goods are going to the port of Rotterdam or the airport in Amsterdam.
Export	Goods that are transported from the Netherlands to foreign countries, most of the times the goods are from the port of Rotterdam or the airport in Amsterdam.
Sea Transport	Transport over deep sea. Mostly import or export transport.
Barge Transport	Transport over inland waters.
Air Transport	Transport using airplanes. Mostly import or export transport.
Terminal	Is a distribution intersection or storage point where vehicles arrive or depart at ports and airports.
Bill of Lading (B/L)	Issued by the shipowner or his agent to a freight forwarder on receipt of goods from shipper agreeing to deliver goods at destination.
Air Waybill (AWB)	Receipt issued by an international airline for goods and an evidence of the contract of carriage.
AIS Data	Transmitted data by ships via radio wave to other ships as well as to inland stations.
ADS Data	Transmitted data by airplanes via radio wave to other airplanes as well as to inland stations and airports.
ETA	Estimated time of arrival.
ETD	Estimated time of departure.

To explain why planners want to update the current status of active shipments an example is given. Take figure 13 where a customer sends a package to the delivery destination. The package is collected by truck and the customer can transport the package by rail or ship. At the arrival at the port of destination or the terminal station, the last part of the shipment is by truck. Focusing on the transport by ship, the planner wants to know whether the package is on the ship and when the ship will arrive at the port of destination because the planner must arrange a truck to transport the package to the delivery destination as well. The truck must be ready to retrieve the package when the ship arrives. Costs for the logistics provider will be higher the moment that a truck is too early and waits for the package. This occurs as well when a truck is too late and the package is delivered too late. A planner needs the ETA of the ship as precise as possible so the truck is ready and available at the right time.

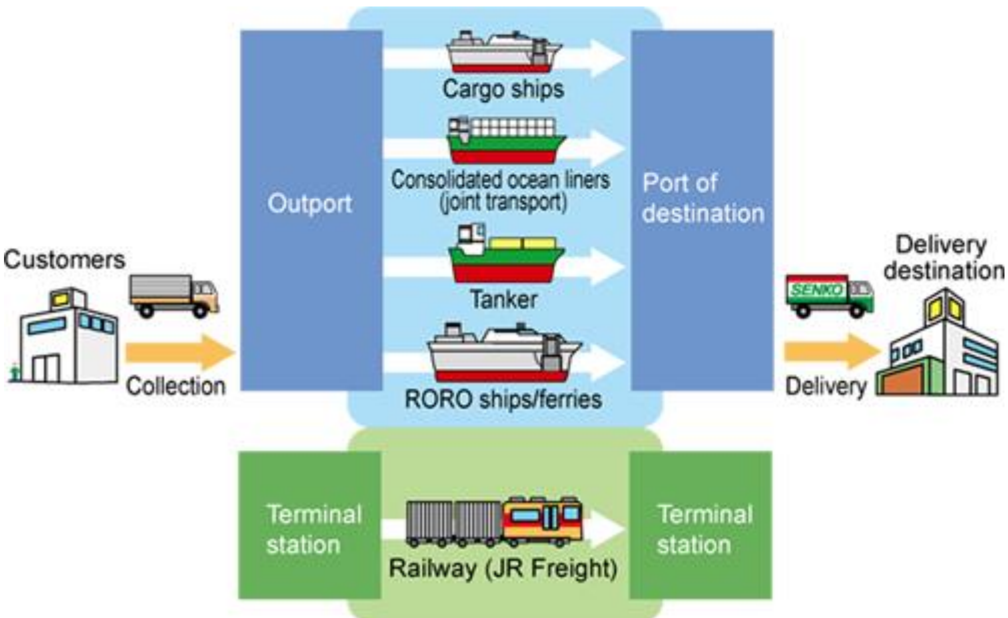

Figure 13: Example scenario of a shipment

3.2. Interview set-up

Interviewing is a method to ask experts about their opinion. Three types of interviews can be defined: structured, semi-structured and unstructured interviews (Bernard, 2006). For the interviews in this chapter a semi-structured interview is chosen. They provide the possibility to interview experts about certain topics that are fixed while still being able to go deeper into certain (relating) topics (Cohen & Crabtree, 2006).

Interviews are set up at HST Groep, Neele-Vat and CTT with a planner, who tracks shipments on a daily basis, and sometimes with a manager of the import/export division. During the interviews, the interviewee will be asked to demonstrate how a shipment is updated while the interviewer observed the actions. The interviews were set up in 7 phases:

1. Introduction
2. Process explanation
3. Important process choices

4. Observation of updating a shipment
5. Required data
6. Data users would like to use
7. Discussion

More detailed information about the interview setup as well as the questions asked can be found in Appendix A. Table 6 shows who is interviewed from what company and what his/her role within the company is. With HST Groep interviews were held with planners with both sea and air freight division while at Neele-Vat interviews were held with the sea freight division. At CTT, a planner is interviewed for barge since they only have inland transportation.

Table 6: List of interviewees

Interviewee no.	Company	Role
1	HST Groep	Planner for sea freight Import Shipments
2	HST Groep	Planner for sea freight Export Shipments
3	HST Groep	Planner for air freight Division.
4	Neele-Vat	Administration at Customs Division
5	Neele-Vat	Planner for sea freight Import Shipments
6	Neele-Vat	Manager sea freight Export Division
7	CTT	Planner for Barge Shipments

The results of the interviews are shown in Appendix A.

3.3. General Findings

This section elaborates on the general findings of the interviews.

At all the companies there is, at least, one session a day, most often in the morning, to update the active shipments. In most cases, planners check a shipment each day. Only with shipments that have a travel time for more than several weeks they sometimes check the shipment every couple of days. All companies mentioned that they would like to check shipments with a much shorter travel time even more often than once a day. The reason they currently cannot do this is because other work tasks will be compromised. This can lead to problems when for example a change of the ETA is recognized too late. They see it as an advantage that an integration platform can track shipments more often when it is necessary.

During a shipment, the shipment can have different statuses. The planners use data of the websites to update their own information in the back office system. Most important information that needs to be up to date in the back office system is the current status of the trip. This information is not only for their own insights where the ship is located or sailing, logistics providers use this information is to communicate with their clients as well. In general, the status of a shipment has different statuses:

1. Pre-booking: This is a pre-booking of an order which is not scheduled.
2. Scheduled: The trip is scheduled for the container. The customer knows when and where the container will be picked up.

3. Departed: This means that the ship has loaded the container and has departed from the departure location.

4. Arrived: The ship has arrived at the arrival location and the container is ready to be picked up.

5. Incident: This status represents that something happened that is not according to schedule. For example, a container is not picked up because the ship was too full and will be picked up later by another ship.

3.3.1. BL and AWB Number

During the observations, the importance of BL and AWB documents are recognized. It is evidence of the contract of carriage for Deep Sea and Air transport. For both BL and AWB, a definition will be given.

A Master B/L (B/L) is issued by the ship his owner or his agent to a freight forwarder on receipt of goods from shipper agreeing to deliver goods at destination. This document is received within three days after the actual departure of the ship and states whether a container is on the ship as scheduled and booked. The logistics providers know only after receiving the Master B/L for sure if the container is on board. A Master B/L is a unique code and is built like:

Master B/L unique code= "SCAC Code" + "Booking Number"

The SCAC code is the Standard Carrier Alpha Code and is a unique two-to-four-letter code used to identify transportation companies (NMFTA, 2015). The SCAC Code identifies the transportation company that does the shipping while the booking number is the booking number of the transport.

The Air WayBill (AWB) is the leading document for tracking the air shipment. The air waybill has similarities with a master B/L. It is a receipt issued by an international airline for goods and an evidence of the contract of carriage. The air waybill (AWB) has a unique number to identify the receipt. The AWB number has 11 digits and has three parts:

Air Waybill number = "Airline Prefix" + "Serial Number of the AWB" + "Check Digit"
3 digits 7 digits 1 digit

The airline prefix identifies the airline. Each airline has been assigned a 3-digit number by IATA, so from the prefix the airline who has issued the document can be identified. The last digit is called the check digit. It is arrived at in the following manner: seven digits running numbers are divided by 7. The remainder becomes the check digit.

3.4. Processes

This section will give the important findings relating to the track and trace processes at the different companies. In the introduction, the situation is sketched and explained why the planners want to track and trace their shipments. For the modalities Deep Sea, Barge and Air, an overview of the track and trace process is discussed as well as what data and sources are necessary for this process. The process flows of the identified processes at the companies can be found in Appendix A.

3.4.1. Deep Sea

Shipments over deep sea can have long travel times. A shipment from China to the Netherlands can take up to a month or more. During these weeks, the shipment will be regularly tracked to follow its progress.

Figure 14 shows the general track process for deep sea shipments. Since the HST Groep en Neele-Vat are mainly focusing on import, figure 14 is based on the import process. Export processes use the same information, only the request of this information might in a different order. In general, the tracking process looks like:

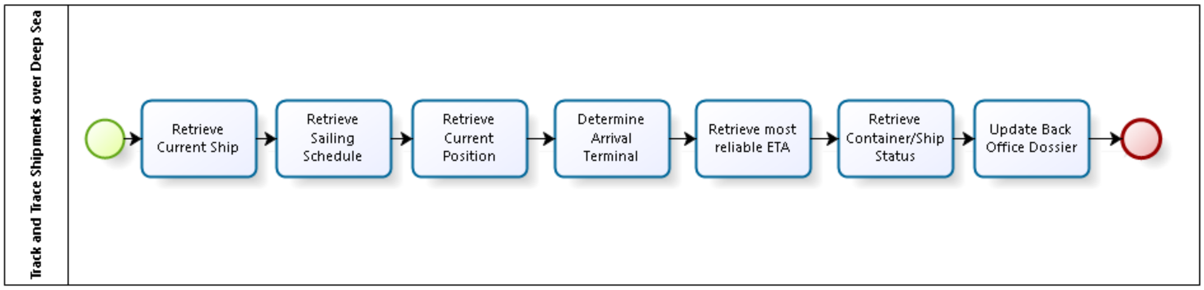

Figure 14: General track and trace process for deep sea

The first step of the process is immediately the most important step in the process. It is the identification of the ship where the container is stored. Based on the ship name, the rest of the information can be retrieved. It is possible that during the trip the container will get transferred to another ship unannounced. Therefore, it is important that the first step is checking at the carrier on what ship the container currently is.

By retrieving the sailing schedule of the ship at the carrier website and its current location at the marinetraffic website (contains AIS data) the planner can check the progress of the ship and the ETA at Rotterdam. When the ship travels to Rotterdam, terminal websites are becoming more reliable and accurate. A ship arrives at a terminal to unload and load containers. During the trip the ship will be signed up at a terminal in Rotterdam or the terminal can change over time. The terminal websites show information as the ETA and the status of the ship and the status of a container.

Based on this information planners can predict the ETA of the ship and update the dossier with the found information as the ETA and ship and container status. Information is retrieved from different sources. Table 7 shows what data is used by planners and where these are retrieved.

Table 7: Data retrieved from website types for deep sea

Data	Website Type	Data	Website Type
Sailing Schedule	Carrier	Ship Name	Carrier
Container Status	Terminal	ETA	Carrier, Terminal, AIS
ETD	Carrier	Ship Status	Carrier, Terminal
Geographical Position	AIS	Course Ship	AIS

The data needed from the data or back office system to retrieve the above data are shown in table 8.

Table 8: Data needed from dossier or back office system for deep sea

Ship Name	B/L Number	Container Number	ETD
ETA			

For Deep Sea relatively many data are publicly available on the internet. A distinction can be made between three different website types: Carrier, Terminal and AIS. Carrier websites are websites belonging to carriers. A carrier is contracted by the logistics provider and maintains its own website where they publish sailing schedules of ships as well as current data about ships and containers. Terminal websites stores data about what happens on the terminal and are maintained by terminals. Examples of terminals in Rotterdam are ECT, APMT 1 and APMT 2. These websites store information as the status of a ship or container. AIS websites are public websites where AIS data are collected and presented. Ships are obligated to send AIS data to inland stations. AIS stands for Automatic Identification System. A clear definition of AIS is given by Wijaya et al. (Wijaya & Nakamura, 2013):

"AIS is a means for ships that make it possible to transmit data via radio wave to other ships as well as to inland stations. The transmitted information can be divided into three categories: (1) Static Information, which includes ship's name, International Maritime Organization (IMO) number, Maritime Mobile Service Identity (MMSI) number, and dimension. (2) Dynamic Information, which contains ship's position, Speed Over Ground (SOG), Course Over Ground(COG), current status, and rate of turn. (3) Voyage-specific Information, which consists of destination, Estimated Time of Arrival (ETA), and draught."

3.4.2. Barge

Barge is the transportation by ship over inland waters and is the main modality for CTT. The transportation time can be less than a day till a week. For barge, less information is publicly available than for Deep Sea. AIS data are richly available just as terminal data at the large ports as Rotterdam and Amsterdam. For smaller inland ports information is not public available. These smaller ports have no interest in sharing data. Therefore are barge processes mainly based on AIS data and the information from larger terminals. The process, again import, is shown in Figure 15.

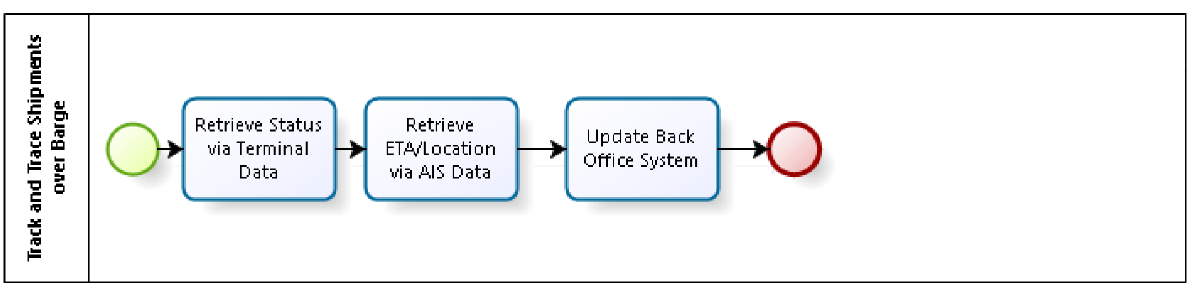

Figure 15: Track and trace process for barge

Before a ship leaves with containers, the deep sea ships are checked every day for the status of the container. This creates the possibility to change containers to different ships when a container is too late or early. The next step is to check the status of the departing ship at the terminal. During the trip AIS data are used to check the current position and ETA.

The data that planners use to update barge shipments are shown in table 9.

Table 9: Data retrieved from web sites type for barge

Data	**Website Type**	**Data**	**Website Type**
Ship Name	Terminal	Ship Status	Terminal
Container Status	Terminal	ETA	AIS
Geographical Position	AIS		

The data needed from the dossier or back office system to retrieve the above data are shown in table 10.

Table 10: Data needed from dossier or back-office system for barge

Ship Name	Container Number	ETA	ETD

3.4.3. Air

Airplanes are used to transport shipments over Air. Transportations by Air have a very short time of travel time. Depending on the distance it can take several hours while if the shipments have a stopover, then the travel time can be extended to one or two days. Figure 16 shows the general process for air shipments. This process is not different for import or export transportation.

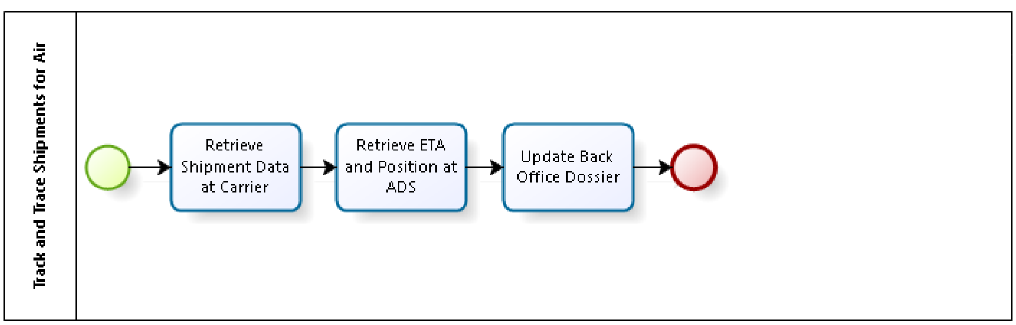

Figure 16: General track and trace process for air

The carrier present on its own website much data about the shipment based on the AWB document number. The carrier stores data about the current status of the shipment, the flight schedule and expected ETD and ETA. Since these websites deliver much shipment details, this is the first type of websites that planners visit. When a flight has departed, the ETA on the carrier website is often less accurate than other sources. During the flight, several ADS sources show current ADS data of airplanes. Such ADS data are used to determine ETA, ETD and possible delay.

The data retrieved from the different websites types to update the dossier is shown in table 11.

Table 11: Data retrieved from web sites type for air

Data	Website Type	Data	Website Type
Flight Schedule	Carrier	Flight Number	Carrier
Container Status	Carrier	ETA	Carrier. ADS
ETD	Carrier, ADS	Flight Status	Carrier
Geographical Position	ADS		

The data needed from the shipment dossier to retrieve the above data are shown in table 12.

Table 12: Data needed from dossier or back-office for air

Flight Number	AWB Number	ETA	ETD
Container Number			

Just as for modality Deep Sea, relatively much information about Air shipments is publicly available on the internet. Much information is presented by the carrier who give full dossier and status information based on the AWB number. For deep-sea this information was dispersed over terminal and carrier websites. ADS websites show, just as AIS website, ADS data that airplanes are obligated to send. Airplanes are located with an Automatic Dependent Surveillance System (ADS) (Drouilhet, Knittel, & Orlando, 1996). This system is built for tracking aircrafts and it uses a GPS navigation receiver for determining the position of an aircraft. The data are received by air traffic control ground stations but also by others who own an ADS-B receiver and use this data to present all the aircrafts on a map.

3.5 Websites and Data Representations

In the last section different type of websites are identified that planners use to track their shipments. This section identifies websites that can be used in the integration platform. Furthermore, the difficulties regarding data representations and formats on the different websites are discussed.

3.5.1. Websites

The websites visited by planners can be categorized into different types. Identified in section 3.3 are four different website types:

1. Carrier – Holds information about the shipment based on the shipment's B/L or AWB number as well as ship specific information as sailing schedules.
2. Terminal – Holds information about the arrival of the departure of a vehicle at the specific terminal, mostly these are deep sea ports.
3. AIS – Holds, collects and presents AIS data of ships. This information is sent by the shipper and contains information as ETA, position and the next port call.
4. ADS – Holds, collects and presents ADS data for airplanes. Airplanes are obligated to send data and include information as position and ETA.

Based on the interviews, observations as well as own research, table 13 shows the main sources used and the carriers where the collaborating logistics providers do most business with.

Table 13: Sources per website type

Website Type	Website	Website Type	Website
Carrier	www.maersk.com (Maersk)		http://ecom.kline.com (K-Line)
	www.afklcargo.com (KLM/Air France)		http://www2.nykline.com/ (Nyk Line)
	www.qrcargo.com (Qatar Airways)	**Terminal**	eservices.ect.nl/ (ECT Terminal)
	www.yml.com.tw (Yang-Ming)		www.apmtrotterdam.nl (APMT 1 Terminal)
	tang.csair.com (China Southern Airlines)	**AIS**	www.marinetraffic.com
	www.turkishcargo.com.tr (Turkish Airlines)	**ADS**	www.flightradar24.com
	cargo.virgin-atlantic.com (Virgin Atlantic Airways)		

To identify other available sources for the website types on the World Wide Web, Google is used as search machine. For both carrier and terminal websites, more sources are available. Each carrier and terminal have their own websites where data can be found. Since these are connected to the terminal and carrier, the source will be identified when the particular carrier is contracted or terminal is visited. For AIS and ADS sources, more sources are available. Currently, LSPs use Marinetraffic for AIS or flightradar24 for ADS. Table 14 shows other found sources.

Table 14: Found sources

Website Type	Website
AIS	www.aishub.net
	www.vesselfinder.com
	shipfinder.co
ADS	www.flightview.com
	planefinder.net
	www.flightaware.com

3.5.2. Difficulties Data Representations

All the websites mentioned in table 13 and 14 are analyzed to identify the type of information is available at each website and how it is represented. In section 2.4.4.3, the literature review conflicts are introduced. Zhang identified three different conflicts: schematic, semantic and data (Zhang, 1994). The analysis confirms that between the websites semantic and data conflicts exists.

Semantic conflicts are caused by the different semantic meanings of the same data. An example of a semantic conflict is different meanings of a date and time. A date and time can be an ETD, ETA, a timestamp etcetera. An ETA can mean the arrival time of the whole shipment or its next stop. When messages are sent over the bus integration system, it must be known what the meaning is of each data so data will be mapped correctly during the transformation of the message.

Another conflict is a data conflict. It occurs when data is represented in different scales, precisions or formats. Again a data and time are taken as example. Multiple websites show the ETA of a ship and the planner that has to decide which ETA is most reliable. However between these websites, the same data can be presented in a different format. As an example, an ETA value is retrieved from both Marinetraffic as Maersk. At Marinetraffic, a date is presented as: 2016-01-01 00:00. For Maersk this is: 01 Jan 2016, 00:00. People understand both representations, but for computers both are not the same date. Computers and applications learn one specific notation for dates and times and each date and time need to have the same notation. The same problems occur for other data. Some websites want a B/L number without prefix or need a container number with the prefix and the numbers apart. It is not desirable that both the CDM and the integration platform works with all possible notations. Again, transformations can foster communication by transforming the data into a format accepted by the website or transforming the results back to the format of the common data model.

3.6. Conclusion

This chapter showed the results of the interviews and observations held at different logistics providers. It answers the following sub-questions:

SQ2a: What data sources are used by Logistics Service Providers?

The data sources used by LSPs can be divided into four website types:

1. Carrier – Holds information about the shipment based on the shipment's B/L or AWB number as well as ship specific information as sailing schedules.
2. Terminal – Holds information about the arrival of the departure of a vehicle at the specific terminal, mostly these are deep sea ports.
3. AIS – Holds, collects and presents AIS data of ships. This information is sent by the shipper and contains information as ETA, position and the next port call.
4. ADS – Holds, collects and presents ADS data for airplanes. Airplanes are obligated to send data and include information as position and ETA.

For each of these types, LSPs use a number of websites. Each carrier and terminal have their own website, but for AIS and AIS, only a few sources exist. Table 13 shows the sources planners mainly use. For AIS and ADS information, LSPs currently use marinetraffic or flightradar24.

SQ2b: Which data sources are available?

A search on the internet identified other sources available. Carrier and terminal websites are excluded since every carrier or terminal has their own website. Table 14 shows for AIS and ADS the available websites.

SQ3: What are the similarities and differences between the different representations of the retrieved data?

The websites mentioned in tables 13 and 14 are analyzed to identify the type of information is available at each website and how it is represented. The results confirm that the identified conflicts by Zhang (Zhang, 1994) between the websites semantic and data conflicts exist.

Semantic conflicts are caused by the different meanings of the same data representation. An example is a data that can have many meanings. In a logistic context, it can be an ETA, ETD or a time stamp. Data conflict occurs when data is represented in different scales, precisions or formats. At almost every website, data is presented differently. It has several notations while meaning the same. As an example, at Marinetraffic, a date is presented as: 2016-01-01 00:00 while for Maersk this is: 01 Jan 2016, 00:00. In the integration platform, the transformations can foster communication by transforming the data into a format accepted by the website and mapping the data correctly so there is no misunderstanding about the semantics.

> *SQ4: What historical and current data should be stored and shared with the rest of the supply chain in order to make it more efficient?*

During the interviews is identified what data is used by the planner and interesting data for other parties in the supply chain. An overview of all the data combined that planners require from websites or from their dossiers are shown in table 15.

Table 15: The identified data that is necessary for the integration platform

Flight Schedule	ETD	Container Number	Flight Status
Container Status	Geographical Position	ETA	Ship Name
Ship Status	Sailing Schedule	Course Ship	
Flight Number	AWB Number	B/L Number	

Chapter 4 The Architecture

Introducing a track-and-tracing service asks for a change at logistics service providers. Processes and information sharing over the supply chain will change. This chapter elaborates the current and the future infrastructure at Logistics Service Providers and how this change should be handled when they implement the integration platform. It is the second chapter in the second step of the ADR. This chapter starts with an introduction. Section 4.2 and 4.3 explains how a supply chain and its architecture can be modeled. The current and future architecture are elaborated in sections 4.4 and 4.5. The differences between these architectures are explained in section 4.6. Section 4.7 concludes the chapter.

4.1 Introducing the supply chain

A supply chain is a system of organizations, people, activities, information and resources involved in moving a product or a service from a supplier to the customer. It includes the transformation of natural resources, raw materials and components into a products that is delivered to the end customer. In logistics the focus is on moving goods from one place to other. Figure 17 gives an example of a general supply chain.

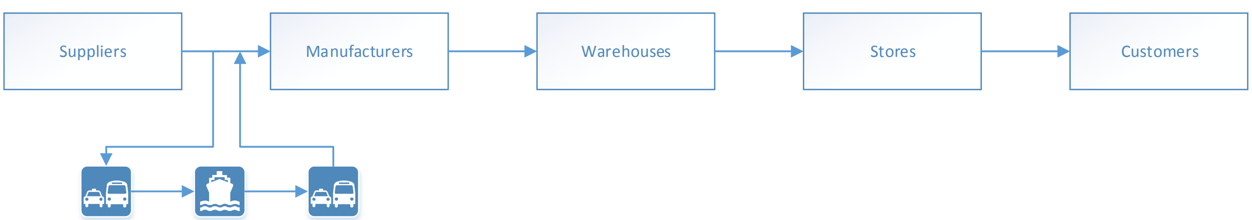

Figure 17: An example of a supply chain

The arrows represent the transportation of goods from one place to the next. The transportation can be divided in different legs. For example, the goods are picked up by a truck at the supplier that brings it to the port. At the port the goods will switch to a different modality and then it will travel by ship to the next port. The goods are picked up again by another truck at the next port that brings the goods to the manufacturer. A LSP coordinates the transportation and contracts different parties that take care of different parts of the transportation. In the example of figure 1, this can be done by three different parties. To coordinate and smoothen the transportation information must be exchanged by all the involved parties. A party in the supply chain must control this information. A LSP is a party that can control and provide such information.

In this thesis the focus of the prototype is to build the tracking of deep-sea and flight shipments. Sections 4.3 and 4.4 presents the current and target supply chain at LSPs with the focus on deep-sea and flight shipments. Section 4.2 introduces the method and tool of how to model and design such architecture.

4.2 Modelling the architecture

In section 2.4.4.1. Archimate is introduced to model architectures and its relationships with the business processes. This section will explain how the architectures of the next sections are designed. The method this chapter follows is derived from Iacob et al. ((M. E. Iacob et al., 2013) where they designed a reference architecture for fuel-based carbon management systems in the logistics industry. The method follows a three step approach.

1. Get insight in the domain of the scope of the project in order to understand what activities are currently performed and where problems arise during the process.
2. Gain insight in the Information System landscape at Neele-VAT, HST Groep and CTT. During this step is identified what systems are used, where the information comes from and what information is necessary.
3. Combine the previous steps and unite the business and Information Systems landscape and bridge the gap between the needs of the business and how the Information Systems supports the business. The result of this step is the architecture of the current and future situation.

The first two steps are performed in the previous chapters. Chapter 2 gained insight into what is written in the literature about the transportation domain and scope. Chapter 3 gained a more practical insights of how it currently works within different companies, what processes they have, which problems arise, what systems they use and how everything relates. This chapter describes the third step of this three step approach: combine the acquired information to create architecture models.

4.3 The current architecture

This section explains the current architecture of the application. The architecture serves to show how business processes relate to the tools and how currently the architecture generally looks like at LSPs. This and the future architecture will be used to show how the architecture, the applications, the business processes and their relations change with the introduction with the tool that has to be designed.

The current architecture shows that in the current situation planners are manually updating the dossiers by visiting online websites for information and updating it in their back-office system while be regularly contacted by a carrier or customer about the status of a specific shipment.

The current architecture is shows in Figure 18 as well as in Appendix C. Each of the three layers are explained in more detail in the next sections.

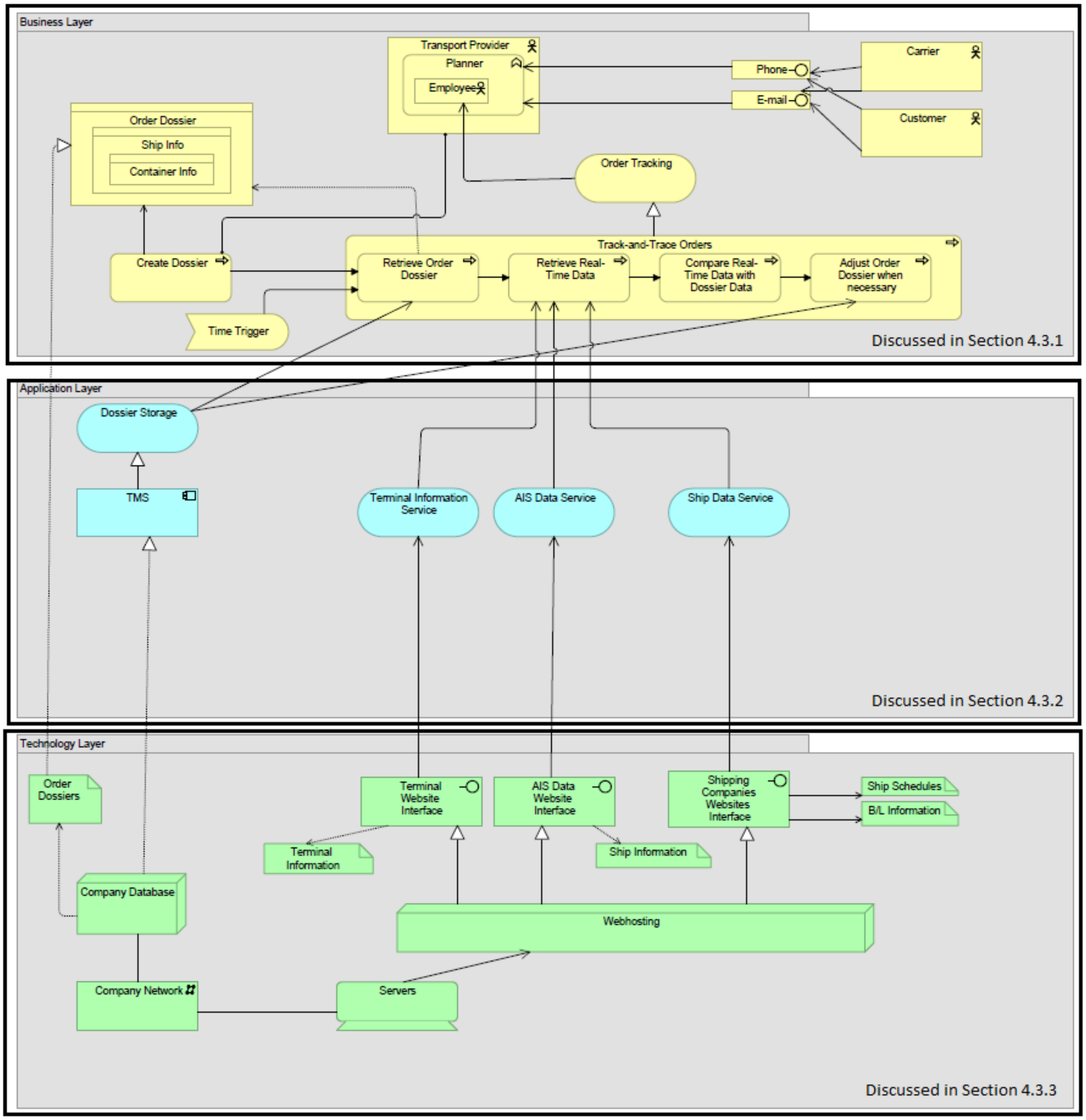

Figure 18: Current architecture

4.3.1. Business layer

The business layer consists out of the top layer of the three group blocks, also shown in Figure 19, and can be identified by the yellow blocks.

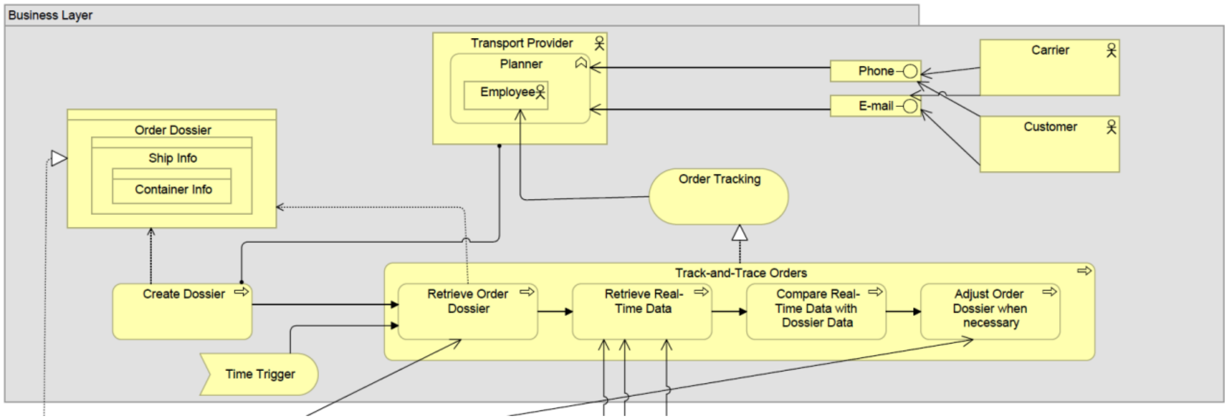

Figure 19: Business layer for the current architecture

The business layer shows the current business process of how organizations currently do their track and tracing process. The business process includes three actors. The main actor in the process is the employee in the role of the planner. The planner is responsible keeping the data of the shipments up to date in the back-office systems. The other two actors are the carrier and the customer. The customer regularly asks by email or phone the status of their shipments while the carrier contact for information about their shipments. The planner can answer these questions with the data that is kept up to date.

The first step of the business process is the creation of an order dossier when the logistics provider and the customer agree on the transport of the goods. A dossier includes details about the shipment, the initial booked vehicle and the container in which the goods are transported. The dossier is assigned to one or multiple planners who are responsible for updating the dossier. On selected times, for example each day at 9 AM, the dossiers must be ordered and the track and trace process is initiated. The track and trace process consists of four steps:

1. The first step is the retrieval of the dossier from the Transport Management System (TMS). The TMS will be explained in more detail in section 4.3.2. From the dossier data is retrieved that is needed as input for the websites.
2. The second step is retrieving of the data from the websites. Manually, websites are opened in the browser and information is retrieved.
3. The third step is the comparison of the data from the websites with the data from the TMS to check if data is still up to date.
4. The last step is updating the dossier in the TMS. If the data is not up to date anymore, the planner decides what information is most reliable and updates the dossier.

4.3.2. Application layer

The application layer consists out of the second layer of the three group blocks, also shown in Figure 20, and can be identified by the blue blocks.

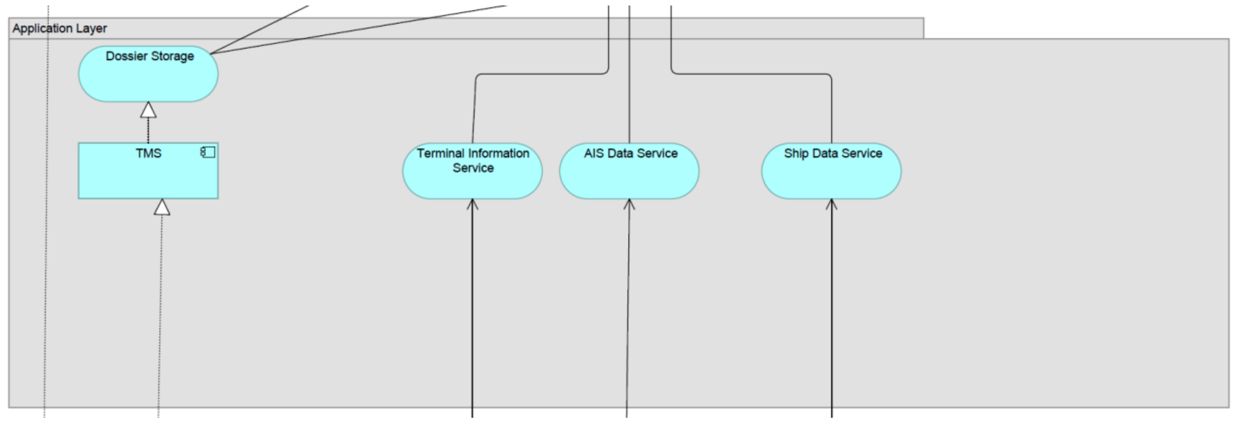

Figure 20: Application layer of the current architecture

In the business layer four services are identified that supports the track and trace process in the business layer. The first service is the dossier storage. By the creation of a dossier the dossier is stored in the TMS. During the track and trace process the dossier is retrieved from the database and updated. The TMS is a software system that is used by logistics providers to manage their transports. In a TMS includes functionalities as order, planning and financial management.

The other three services are the services that the websites deliver for the planner. In chapter 3 are four website types are identified. The terminal information service (Terminal website) offers terminal specific information. The AIS data service (AIS and ADS websites) offers real-time data of the vehicles. The ship data service (Carrier websites) offers carrier specific information as the container status and the vehicle schedule.

4.3.3. Technology layer

The application layer consists out of the last layer of the three group blocks, also shown in Figure 21, and can be identified by the green blocks.

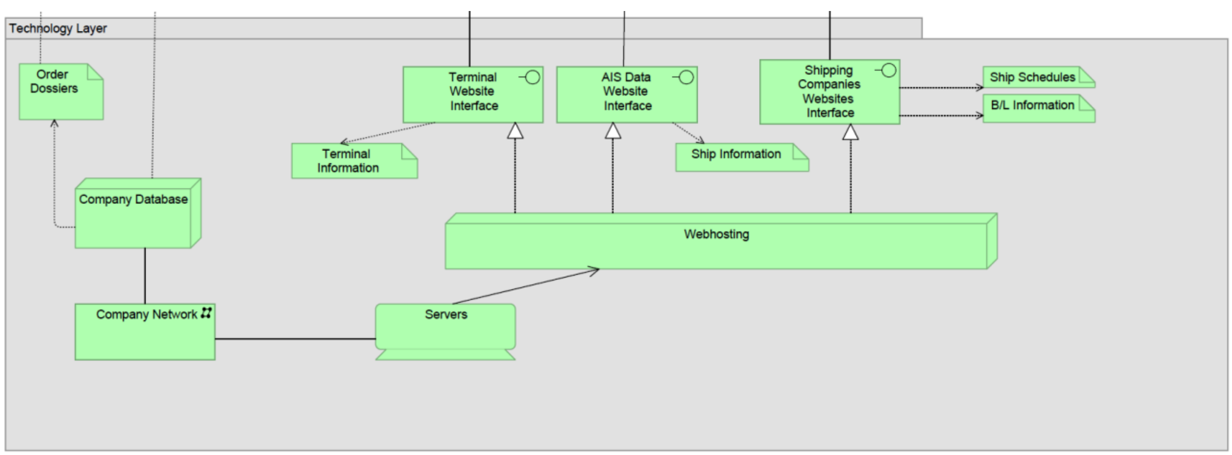

Figure 21: Technology layer for the current architecture

In the previous layers is discussed that dossiers are stored in the TMS but on a technology level, the TMS is connected to a database which stores all the (historical) data of the company. The database runs on internal servers from the logistic company making sure that that the data is not openly available for others. In the application layer are the three services discussed of the websites. The services are each connected with an interface. The planner communicates with this interface and based on the actions of the planner the interface shows the results. All the websites store information in their own databases and are hosted by a webhosting service. The webhosting service has multiple servers on which the websites are running.

4.3.4. Analysis of the current architecture

The current architecture does not only show the current situation for the planners but also visualizes the problem of the logistics providers and the planners. In the business layer, the manual process for the planners is drawn. According to the interviews, this process takes, per planner, half an hour on average. The connections via the services to the website interface shows that planners directly communication with the websites. Besides this process, customers and clients currently requesting information about their shipments by email and phone. Customers and clients are mainly requesting update statuses. Both the updating process as the communication with customers and clients are time consuming.

The technology layer shows no connections with other parties in the supply chain. The required communication is done by phone or email but no information is directly shared since the parties are not integrated. If a party must be connected to the system, this can be either a tightly coupled one-to-one connection or a loosely coupled connection via an integration bus.

In the target architecture, these problems will be checked off and solved.

4.4 The target architecture

In this subsection the target architecture is discussed. The target architecture is the architecture after the implementation of the integration platform. The architecture will be useful to understand the impact of the platform on both the organization's business processes and their IT landscape.

In the target architecture the integration platform is implemented. A planner does not communicate with all the websites apart but communicates with one portal, hosted by the application in the application layer, which shows all the information that is retrieved from the websites in one format. The application triggers automatically the updating process and updates the shipment based on the results. It also means that within the application the time trigger can be different than in the current architecture. Currently the track and trace process is mostly once a day, the application can trigger this process more often.

The target architecture is shown in Figure 22 as well as in Appendix C.

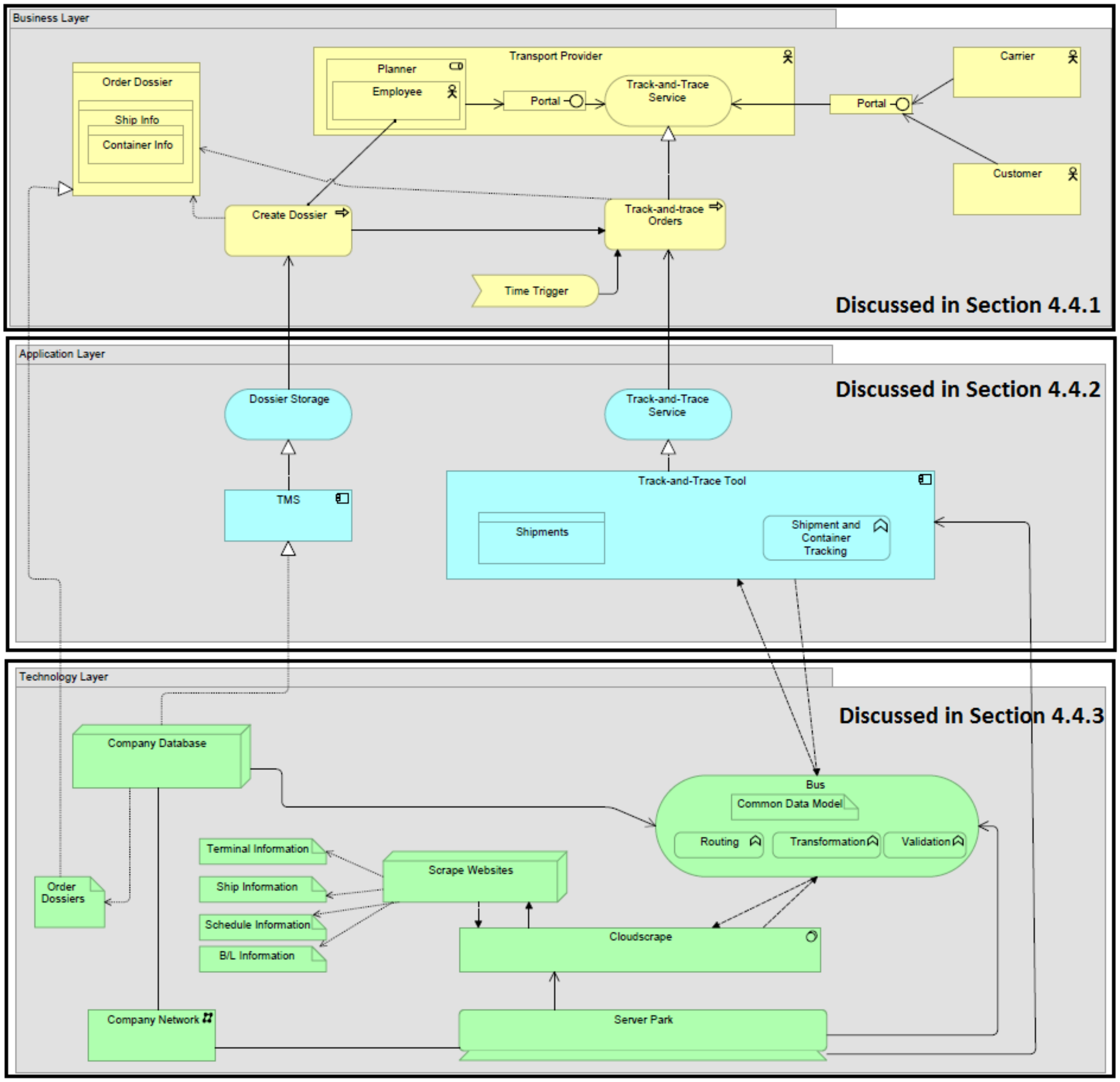

Figure 22: Target architecture

4.4.1 Business layer

The business layer contains still three actors and is presented in figure 23. While the planner has still his responsibilities to create the dossier in the TMS and check the status of the dossiers, the planner communicates in the target architecture with a portal for the status of the shipment. The portal shows all the up to date information the planner needs to check the progress. The same portal is used to communicate with the carrier and the customer. Instead of the communication by phone and email, the portal delivers this same information to the actors any time.

The portal is the interface of the track and trace service and displays the information that the track and trace service delivers. During set times the track and trace process is initiated and all the shipments are updated and shown to the actors via the portal.

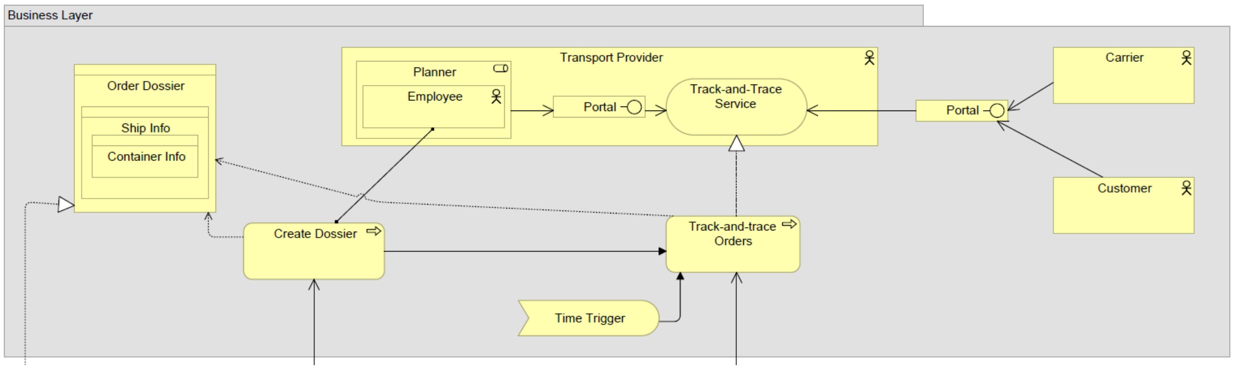

Figure 23: Business layer for the target architecture

4.4.2. Application layer

The application layers, shown in figure 24, consists of two services that support the activities in the business layer. As in the current architecture the dossier storage service stores the dossier during its creation in the TMS. The track and trace service supports the track and trace orders activity in the business layer. The service is connected to a tool/application that is responsible for the automatically retrieval of the information on the websites and the business logic to process the information. The application stores the dossiers in its own database as well and keeps them in this database updated. The application also makes sure that the information in the portal is shown unambiguously to the user.

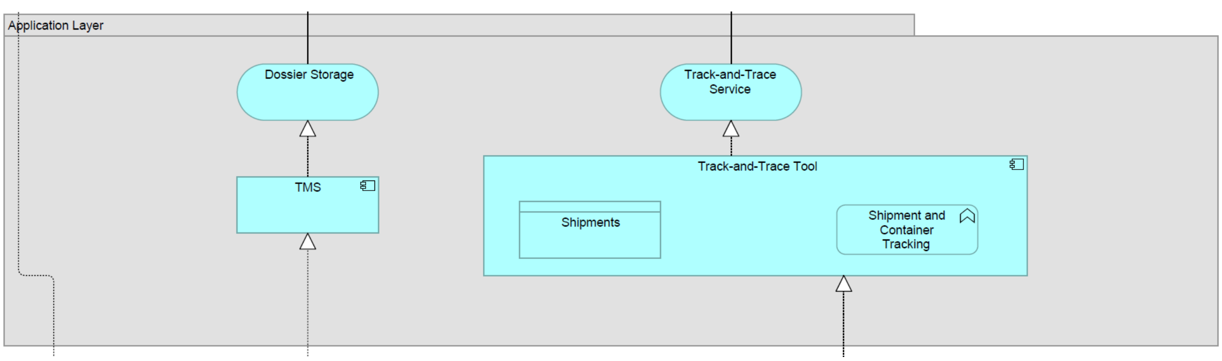

Figure 24: Application layer for the target architecture

4.4.3. Technology layer

In the technology layer, shown in figure 25, the order dossiers are, just as in the current architecture, stored in the company database that is hosted by the company network. In the application layer an application is added that needs to communicate with other systems in order to retrieve data. The integration bus regulates this communication. The integration bus is responsible for the routing, validation and transformation of the messages and is connected to the application in the application layer, the company database and Cloudscrape, the web scraping tool. From the company database new created dossiers are retrieved and send to the application to store it in that application. The bus communicates with Cloudscrape to for the retrieval of information from websites. Clouodscrape scrapes the selected websites that stores the information.

The company database, Cloudscrape, the integration bus as well as the application run of different servers.

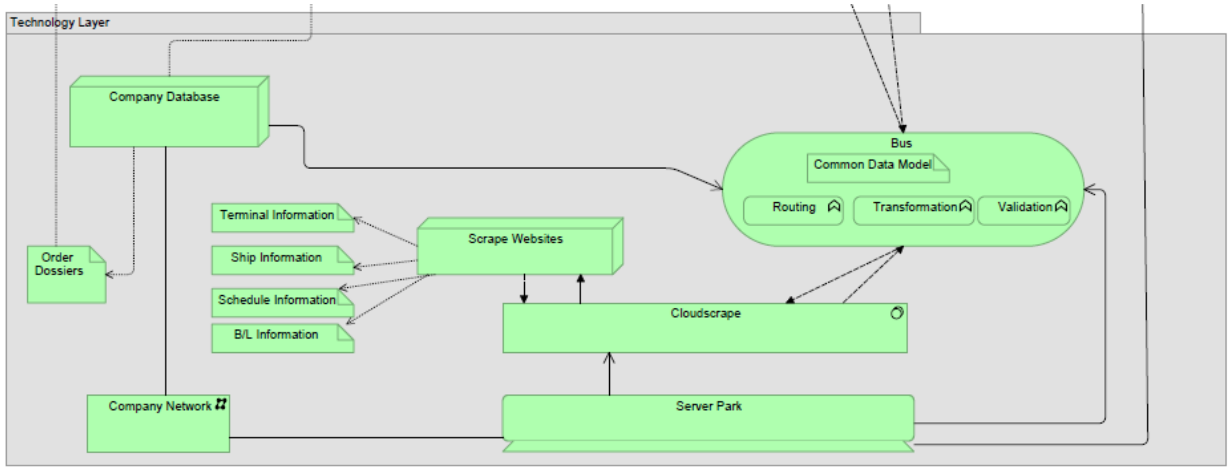

Figure 25: Technology layer for the target architecture

4.5 Differences

To define the impact of the implementation of the integration platform, both architectures must be compared. A gap analysis represents the differences between two architectures and will show the impact. This section will discuss the differences between two architectures and how this influences the business process. The gap analysis is shown in figure 26.

The implementation of the integration platform and portal changes the business process of updating the shipment. First of all, communicating with the back-office systems during the updating process is removed and will go via the portal. This has multiple advantages.

1. The planner communicates with the portal which shows the latest information about the shipment. This will reduce the time that the planner spends on updating the shipments and intervening when necessary as well as reducing the complexity of the process for the planner. The status is kept updated in the application instead of the back-office system that in this architecture serves to store the orders and send them to the application. Via the bus, the application can send status updates to the back-office system and update the back-office system with the same information as in the application.
2. The application and portal can show and store more information than in the TMS. The TMS is a large software package and is stores only the necessary data for the shipment. The application enables planners to request additional data and connect shipment to each other to increase the ability to be synchromodal.
3. The portal can send automatically messages and warnings to the planner based on the retrieved information. Ideally, the planner only have to access the platform when a message or warning is received.

4. The planner will communicate with one source, the portal, instead of multiple sources. This reduces the chance on misunderstanding the information on the websites. In the portal all the data, for each modality, is shown in the same format.
5. The data can be updated more often in the portal than it is currently done manually. Based on the preferences of logistics providers the frequency to update a shipment can be higher resulting in more up to date data during the day.
6. The data stored in the portal can be presented in different lay outs instead of the single lay out in the TMS. Different dashboards can be designed based on the preferences of the logistics providers.

Secondly the portal cuts the direct communication lines with the carriers and customers. Carriers and customers can connect to the platform and receive information via their own interface in the portal. This has two advantages. Firstly, the customer and carrier are kept up to date during the whole process. Secondly, this is an extra additional service which will increase the collaboration satisfaction with both the carrier and customer.

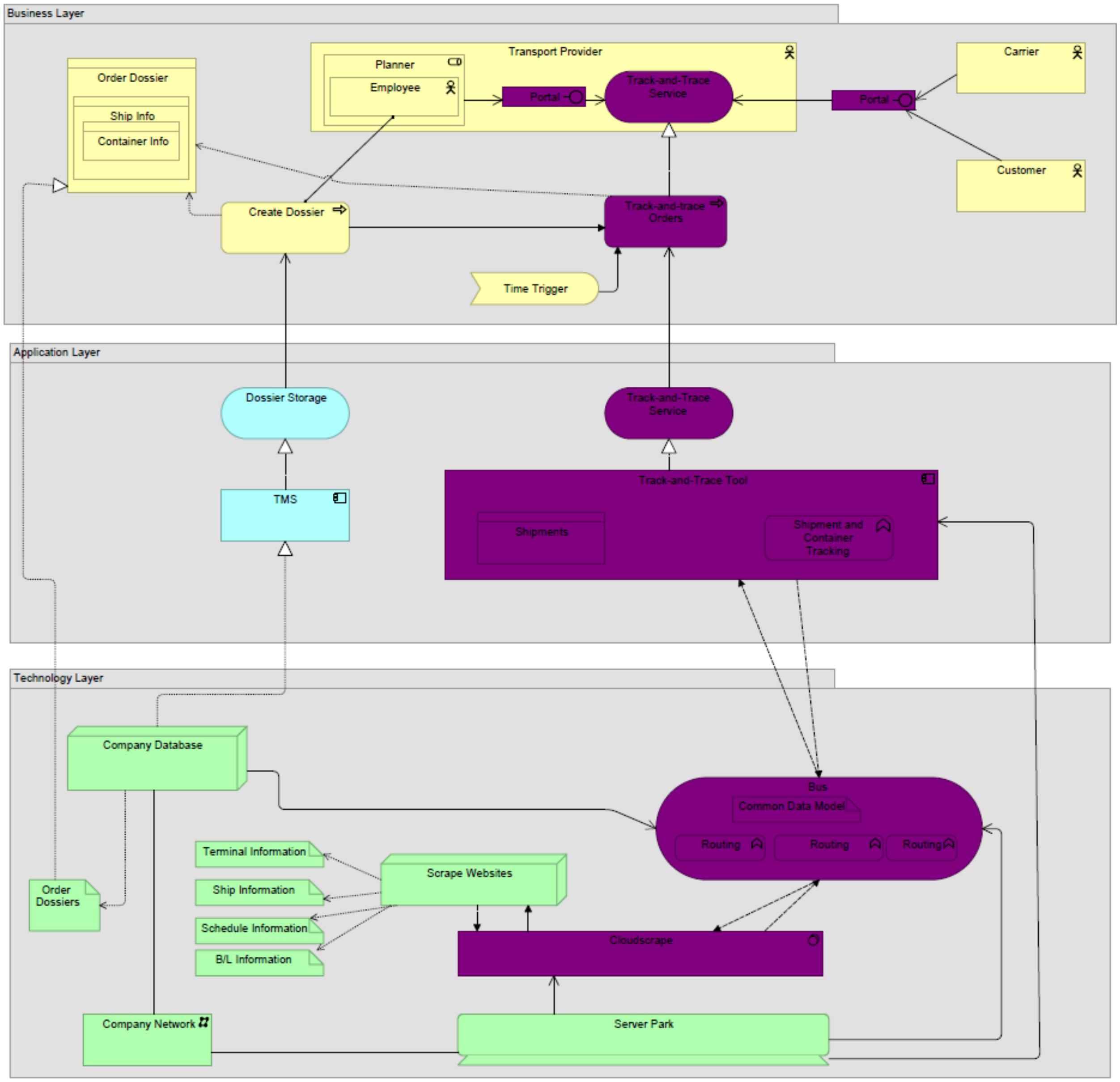

Figure 26: Gap analysis

In the technology layer the integration bus is implemented and connections are built with several systems as the application, Cloudscrape and the company database. In the integration bus the common data model is central. The usage of the integration bus has several advantages:

1. The usage of the integration bus makes it easier to connect other parties as carriers and customers to the platform and with the organization.
2. It stimulates to share more data among the supply chain. More data is available and supply chain partners can easily be connected to the platform.
3. Currently the shipments are manually updated in the TMS. The application stores the up to date information of the shipments. The integration bus makes it possible to automatically update the TMS system with status messages so the shipment in the TMS is updated as well.

4. Based on Bishr's levels of interoperability there is no interoperability in the current architecture. There is no direct communication between the website and the back-office systems. The planners manually retrieve the information, process it and inputs the result in the back-office system. With the implementation of the integration platform in the target architecture the interoperability between systems will be rise to level 5 where the connected systems are provided with a single global modal (common data model) that is an abstraction of all the underlying remote databases (data models of the systems) (Bishr, 1998). Bishr mentioned that the systems can communicate seamlessly, knowledge on their semantics is still required to reach interoperability. Transformations in the integration bus creates this interoperability by transforming messages from and into a common data model.

4.6 Conclusion

From a synchromodal perspective, the architecture of a supply chain looks like Figure 23, currently. In the future, it could look like Figure 26, including an integrated platform. This answers the fourth research question, which is:

SQ5: What does an architecture of a supply chain look like from a synchromodal perspective?

The chapter focused on the architecture of both the current and the target after the implementation of the integration platform. The current architecture is based on the findings and interviews at the logistics providers. The target architecture sketches the architecture with the implementation of the platform.

The implementation has several advantages for the logistics provider. Automatically track and tracing shipments not only saves time by removing manual actions, it also increases communication with carriers, customers and other parties in the supply chain. The application enables users to store more detailed information about the shipment than the TMS allows and show this information in different dashboards. Linking shipments and the increase of communication among the supply chain increases the ability to be synchromodal as well.

Chapter 5 Design the Common Data Model for Synchromodal Logistics

Enabling efficient communication between different organizations and data sources requires a common data model. The objective of this chapter is to design a CDM for planned and actual information, orders and status/disruptions and answers sub-question 5. The integration bus transforms incoming messages into the CDM before transforming it into a message that the receiving system can interpret and use. This chapter introduces the design of the CDM and is the third chapter of the second step of the ADR. The first section introduces the chapter. Section 5.2 identifies what kind of data types will be stored into the CDM. In section 5.3 the data are identified that is necessary for the CDM. The CDM will be designed in section 5.4. Section 5.5 concludes the chapter.

5.1 Introduction

In integration processes, integration turns many small applications into one large application (Britton & Peter, 2004). Integration enables communication between applications. Each (back office) system expects or sends a message in its own format. The CDM plays an important part because the CDM is an uniform format where all the separate formats can be transformed into. In this thesis, back office systems are at LSPs that contains order dossiers.

Chapter 3 showed the different processes that organizations use to track their shipments. For these processes, the used and necessary data are identified. During the interviews is identified that the current processes and methods are not sufficient. It is time-consuming and important changes in the status of the shipment are sometimes not identified in time. The CDM will be designed based on the information needs of the LSPs. This includes that the processes, the data sources and data types with the needs of the organizations will be combined. This process is shown in Figure 27.

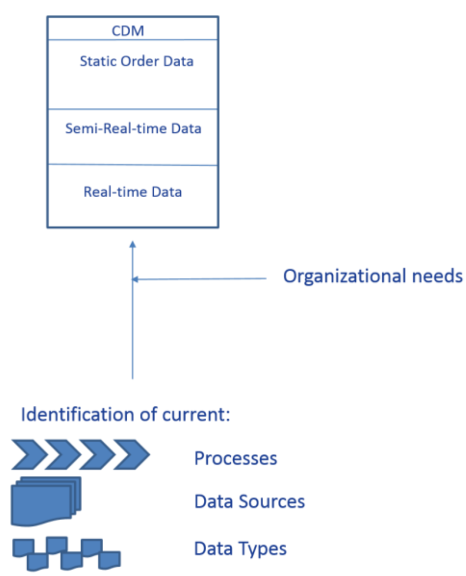

Figure 27: Overview of the CDM design process

The first step is the identification of what the current processes of LSPs are and what information they currently need and have to trace their shipments. The second step is the additional information the organization wants, but they currently cannot do due different reasons. Both steps are combined to a Common Data Model that can store up to date information for shipments over deep sea, air or barge. The current processes and the identification of important information are described in chapter 3. In the next sections, the design of the CDM will be discussed.

5.2 What data is covered in the common data model?

This section discusses how the CDM is built. From interviews and observations, as described in chapter 3, can be derived that there are three types of information. The first type is the static order data. This information is retrieved via the back office system. This information consists of basic order data as from and where to has the shipment to go, who is the shipping company, in what container are the goods transported and who is within the LSP the owner of this dossier. This data are rarely changed and are therefore almost constant during the whole shipment. This is also shown in figure 28 as the information that defines the trip from A to B.

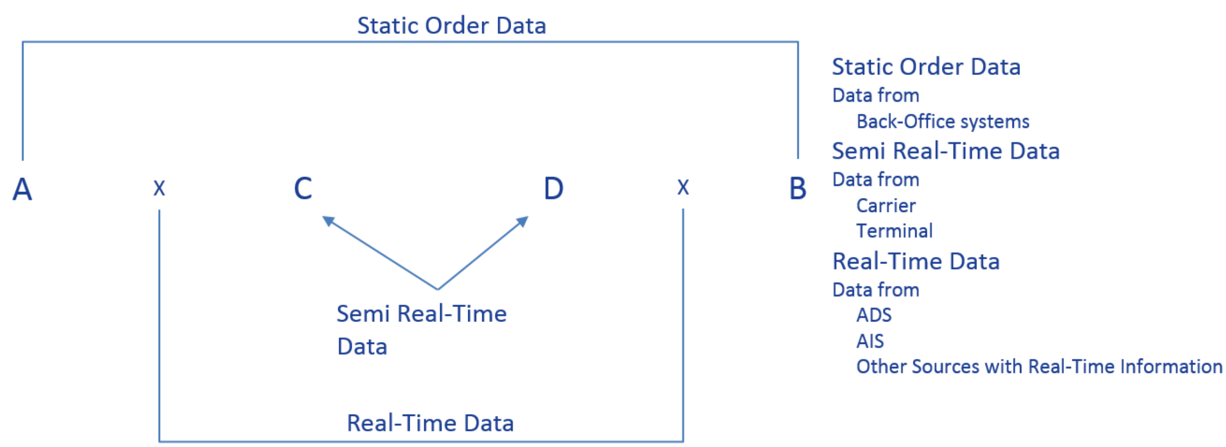

Figure 28: The data that the CDM consists of

The second part of the CDM is the Semi Real-Time Data. This includes data that are regularly updated but changes only once or twice a day. This data include data as (updated) vehicle schedule, port calls and current vehicle and is mainly retrieved from the website types carrier and terminal. This means, as shown in figure 28, that calls in C and D are also known. The third part is the real-time data and is mainly data from website types ADS and AIS. This includes information as the current location information, weather information, traffic information or disruptions. This is shown in figure 28 as the real-time location of a vehicle between A and C or D and B.

Currently most LSPs use only the first two parts of the CDM and some LSPs use some real-time data. This real-time data are often the location of the vehicle that is tracked. Organizations have needs to include more real-time data in order to make better and more efficient decisions. However currently this cannot be done because the current process is already time-consuming. Real—time data can consist of weather information, traffic information, and etcetera.

5.3. The Common Data Model

In this section, the common data model will be designed following a bottom-up approach meaning that the data model will be designed using information collected from practice. This approach will ensure that the common data model will be as compact as possible. Using industry standard messages are too large and extensive and use too many data that is not required for the goal of the common data model. The section follows the method of Böhmer et al. (Böhmer et al., 2015). The method, introduced shortly in chapter 2, consists of four steps:

1. The first step is the problem and requirement definition. It explains the needs for a common data model in the logistic sector for track and tracing and explained in Chapter 1.
2. The second step consists of the analysis of the considered domain and is done in Chapter 3 and 4. In Chapter 3, interviews are held to identify the processes, the different scenarios in these processes and what sources and data are used. In Chapter 4, the architecture designed and showed how the data model fit in the architectural view and effects the organization when implementing the integration platform.
3. The third step is the design of the data model and explained in this chapter.
4. The last step, maintenance, is done by validating the data model and revise it when necessary.

The common data model will be designed according with the following requirement that the common data models holds all the information that logistics providers require to update their shipments.

In chapter 3, the necessary data are identified that are required to retrieve the data. All the data will be combined and all duplicates removed. This leaves a set of unstructured data which will be the base of the CDM. The next subsections sum what data are necessary for track and tracing.

5.3.1. Deep Sea and Barge Transport

The scope of the thesis defines that the prototype includes the modalities deep sea, barge and air transportation. This subsection identifies what data are necessary for deep sea and barge transport. These modalities are grouped together because both modalities use ships as vehicle and, therefore, both modalities have a lot in common.

To track a deep sea or barge trip three types of information is necessary as shown in subsection 4.2: Static order data, semi-real-time order data and real-time order data. The static data define the order and will be used as input as search terms. Since goods can be transferred to other ships, the only static information to track goods are the B/L number, the Container Number, in which the goods are transported in, where the container should depart and arrive and who ships the container. The order data define also the customer of the shipment.

The semi real-time order data contain data as the status of the ship, ETA, ETD and the status of the container. This B/L number and container number are used to find on what ship the container is currently placed, its status and what its ETA and ETD is. The ship name can be used to retrieve sail schedules so the progress of the shipment can be measured.

The real-time order data are the data that contain the geographical position of the ship. This is stored as a latitude and a longitude coordinate and a course. Table 16 shows what data are necessary to be able to track shipments for Sea and Barge.

Table 16: Data necessary for modalities sea and barge

Data for Sea and Barge		
Ship Name		
Ship Status		
From		
To		
ETA		
ETD		
Carrier Name		
Customer		
Container Number		
Container Status		
Departure Location		
Arrival Location		
B/L Number		From
Sail Schedule	-------	To
Latitude		ETA
Longitude		ETD
Course Ship		Status

5.3.2. Air Transport

Air transport is less complex to track and that is mainly because of the limited websites that are necessary to retrieve the required information. There are commonalities with the Sea and Barge transport. The static order data contain the basic order information as the departure and arrival location of the shipment, the planned ETD and ETA, in what container the shipment is and the AWB number. Semi Real-Time order data contain the flight schedule and with what airplanes the shipment is shipped. Also, the status is updated changes during the day. Real-Time data contain the location of the airplane.

In table 17 this information is combined with the table of sea and barge to create the data for these three modalities. The underscored data are changed or added to the table.

Table 17: Data necessary for modalities sea, barge and air

Data for Sea, Barge & Air		
Ship Name		
<u>Vehicle Status (Changed)</u>		
<u>Flight Number</u>		
From		
To		
ETA		
ETD		
Carrier Name		
Customer		
Container Number		
Container Status		
Departure Location		
Arrival Location		
<u>AWB Number</u>		
B/L Number		From
<u>Trip Schedule (Changed)</u>	-------	To
Latitude		ETA
Longitude		ETD
<u>Course Vehicle (changed)</u>		Status

5.3.4 Additional Fields

Table 17 shows what data are known on order specific level however more data are needed in order to organize all the order dossiers. In this subsection, these additional fields are identified and added to the CDM.

First of all, it is necessary to relate orders (from now Shipments) to a company and an owner. If this CDM is used in an application that allows multiple organizations to make use of the same application, there is a need to identify what shipment belongs to the right organization. Furthermore, within the organization departments and employees are owners of the shipments. Import and Export departments handle different kinds of shipments and have no need to see the shipments of other departments. This is the same at an individual level.

It is important to add a variable Creation Data into the CDM for search and ordering possibilities. The creation data show when the shipment is created within the application. More important is the type of the shipment. A shipment can be either Import, Export or Inland. Each type has different processes and the type can be used to identify what processes should be used by the application. The same applies to the type of the shipment, in other words, the modality of the shipment. As can be seen in the previous chapter, the processes and data sources between modalities differ.

Other important fields that are added are the Shipment ID and the Leg ID. For extension purposes, this distinction is made. A shipment is seldom only one modality. Therefore, there is a need to include multiple legs (which are all of one modality) to trace the shipment during the whole trip. The Shipment ID identifies the shipment while the Leg ID identifies the leg. It is possible that the leg belongs to more shipments while the shipment has more legs. It is for example possible that a ship has containers for multiple shipments which mean that these containers have the same leg. A shipment can also have its own status, ETA, ETD etc. Other ID's that are added are the vehicle ID, Customer ID and Carrier ID. For a vehicle, a Voyage Number is added as well to store a trip number.

A time stamp is necessary for combination with the geographical position. This enables to track the shipment over time and can be used later for data mining purposes. Table 18 shows the data with the additional information added.

5.3.5. Disruption and messages

To create a feature for the planner to instantly make additional changes to the planning if disruptions along the way happen some additions have to be made to the model. There are all kind of disruptions. Examples of disruptions are lock and tides information or weather information. To show that the model can work with disruption one disruption will be shown as an example. This example will be the weather information for air transport, specifically, weather information at the destination airport. It happens frequently that airplanes have to fly to other airports than their destination airport because of the weather at the destination airport. To measure this weather information will be stored. If a disruption is identified message should be identified and send to the owner of the shipment. Messages should also be sent if there are important changes in the shipment, for example, the container has arrived at his destination or there is more than a day delay in comparison with the last stored ETA. In all these situations a message is sent to the owner to inform the planner. If necessary, the planner can take action.

Table 18: Data necessary for modalities sea, barge and air with additional fields

Data for Sea, Barge and Air with additional fields		
Company ID		
Company Name		
Department ID		
Department Name		
Employee ID		
Employee Name		
Shipment ID		
Order Type		
Ship Name		
Shipment Status		
Customer ID		
Customer Name		
Vehicle Status		
Leg ID		
Leg Status		
Carrier ID		
Carrier Name		
Vehicle ID		
Voyage Number		
Flight Number		
From		
To		
ETA		
ETD		
Container Number		
Container Status		
Departure Location		
Arrival Location		
AWB Number		
B/L Number		From
Vehicle Schedule	-------	To
Latitude		ETA
Longitude		ETD
Course Vehicle		Status
Timestamp		

5.4 Design of the CDM

In section 5.3 all the data are identified and listed in table 18. The data listed needs structure in order to create a model. The next step in designing the common data model is to create entities with their attributes, which are listed in table 19, and their relations between the entities. In figure 29 the result presented in a graphical format. The common data model shows how the data is stored and the three parts of the model identified in section 5.2. Based on these three parts the model will be explained.

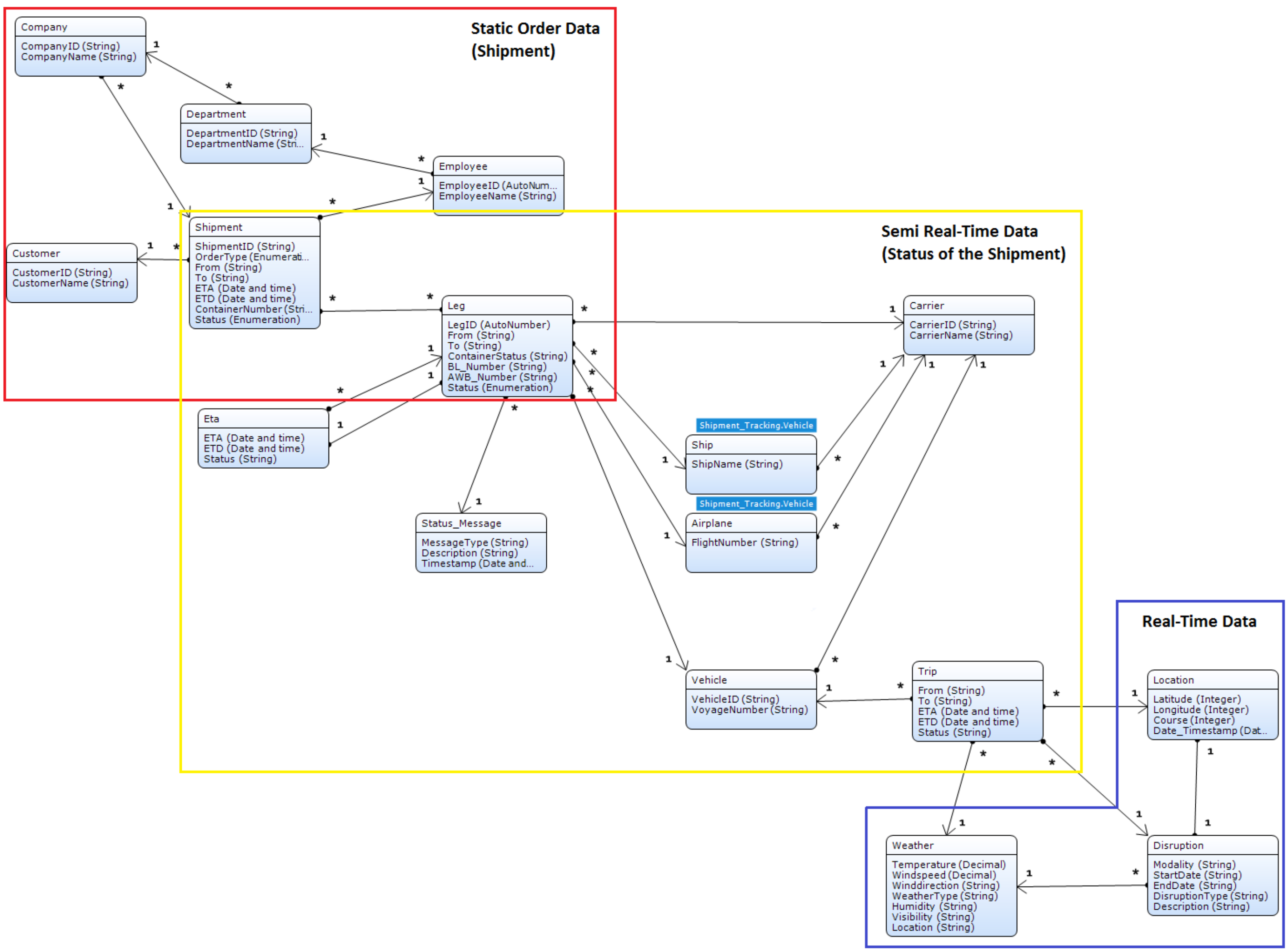

Figure 29: The designed common data model presented in Mendix

Table 19: Description of the different entities of the CDM

Entity	Description
Company	This is the overall entity of the CDM. It defines the Logistics provider that uses the application.
Department	A Company is divided into different departments, handling different types of orders.
Employee	An Employee belongs to a department of a Company and is the owner of his/her dossiers.
Shipment	The entity Shipment defines the order dossier. In the Shipment entity, the shipment of goods from the A to B are defined and the attributes are defined that will not change during different legs.
Customer	The Logistics provider (Company) provides the service of shipping goods for a Customer.
Leg	A Shipment consists of one or multiple legs. A leg is defined by the transport over one modality. A Shipment can have multiple legs but then the difference is that the leg before and after have a different modality.
Carrier	A leg is performed by a Carrier that is hired to transport the goods.
Vehicle	The Carrier transports the goods using a Vehicle.
Ship	Is a generalization of Vehicle for modality Sea and Barge.
Airplane	Is a generalization of Vehicle for modality Air.
Trip	During a Leg, a Vehicle has a certain schedule. The schedule is defined in Trip where Trip is the trip between stop A and B on the schedule.
Location	During the Trip, the location of the Vehicle is retrieved. The Location is stored for every Trip, so the movement of the Vehicle between every Trip can be tracked.
Weather	Weather information is stored per Trip.
Disruption	A Disruption is given when certain values cross set borders. This generates a warning to the owner of the Shipment.
Status_Message	A Status Message is given when important statuses are changed during the Shipment. This generates a message to the owner of the Shipment.

In the model, the static order data is the most important data. This specific part of the CDM is shown in figure 30. These data are received from back-office systems and defines the order in the platform. The order needs shipment-specific information as the departure and arrival location and times, modality and container number as well as information about the customer, carrier and the owner of the dossier to create a basis of the order in the system. Based on this information, the order can be linked to the customer (and her portal) and contains the information necessary to start the automatic information retrieval.

The company is the owner of the shipment (an order) and the highest entity of the data model. A company consists of multiple departments and those departments have multiple employees. This makes it possible to filter the shipments in the platform on different levels, from company specific to an employee specific overview. Shipments have a customer, so shipments can be linked to this customer. This creates the possibility to build customer portal offering customer specific information which leads to less email and phone communication about the status of customer's shipments. A shipment contains, at least, one leg and each leg is the transport with one modality. In a leg is specific information about this specific part of the shipment stored. Using a shipment with different legs, with each a different modality, can lead to an increased synchromodal behavior since the information for an active or current leg influences the next leg on its departure time and location etcetera. The leg is shipped by a carrier which takes care of the transport of the container or goods.

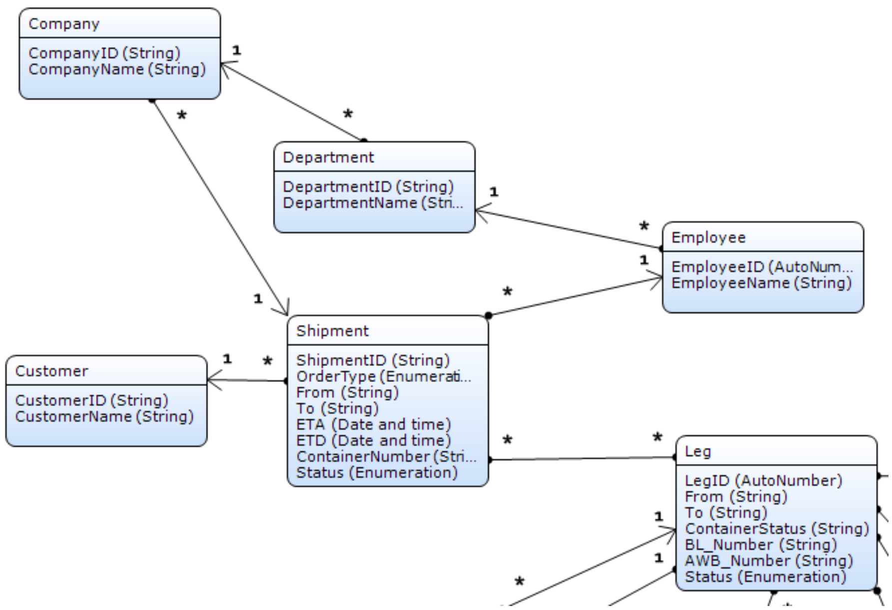

Figure 30: Static order data of the CDM

The second part of the common data model is the semi-real-time data and covers the leg specific information. The semi real-time data is shown in Figure 31. This information is mainly retrieved from carrier and terminal websites, which update their data a few times a day, or is updated based on the information retrieved on a trip level. As mentioned before, for each leg a carrier is hired for the transportation of the container but the actual transportation is by a vehicle which is owned by a carrier. A vehicle has multiple specializations for every vehicle type (ship or airplane). A specialization is necessary for identification of the vehicle since both types are identified with different data. The progress of the leg can be tracked by a vehicle schedule which is modeled as multiple trips. A trip defines the stop the vehicle makes during the leg. The ETA entity stores the retrieved arrival and departure times and stores the optimal ETA for the leg based on the most reliable information. The Status Message that is connected to the Shipment uses Shipment info to inform the planner about the status of the Shipment, for example if the Container is on land.

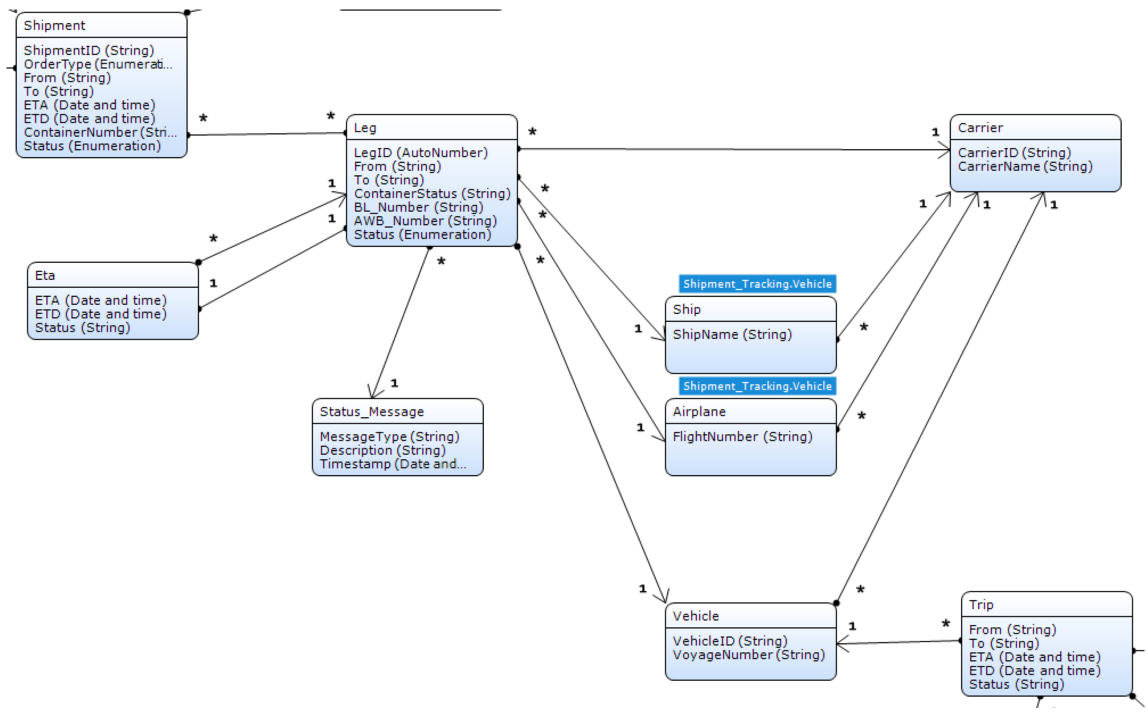

Figure 31: Semi real-time order data of the CDM

The last part of the common data model is the real-time data and is shown in figure 32. It covers the data that is often updated on websites and is mainly retrieved from ADS or AIS websites. This information is linked to a trip which is always the active trip. This way it offers the possibility to do data mining on both a leg and trip level. Information that is linked to a trip consists of the geographical position of a vehicle as well as the different disruptions that are or can be added to the date model.

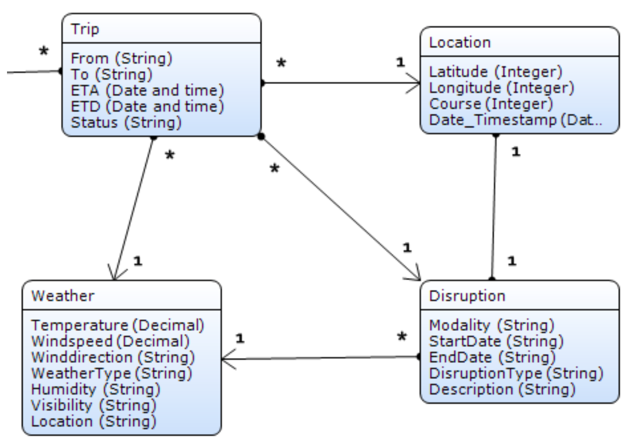

Figure 32: Real-time data of the CDM including the Trip entity

Following the schedule, the progress of the vehicle can be measured. The geographical position of the ship is as well defined per part of the vehicle schedule. The geographical position is used to create visibility to the user and the identification of bottlenecks (via data mining). Disruptions are real-time information that can help the user make better decisions. Disruptions can be weather information, lock times etcetera.

For this data model, the disruption weather information is added. The disruption entity is connected to the weather entity and the location. During the identification of a disruption, a standard message can be generated based on the data as weather information. The location is added so that the disruption message also can send a location to the planner of where the disruption is discovered.

5.5 Conclusion

The chapter answered the first part of the sixth sub-question, the design on the common data model:

> *SQ6: How is a common data model for planned and actual information, orders and status/disruptions designed?*

Three types of data are identified: Static order data, Semi real-time data and Real-time data based on the frequency that the data are updated. The static order data defines the order and the information is mainly retrieved from the back-office systems. This information is mostly static and only a few attributes change during the whole shipment. Semi-real-time data changes a few times a day and is related to a specific leg. Real-time data changes regularly and is related to an active trip.

The design of the common data model took a bottom-up approach using the information retrieved from chapter 3 and follows the method of Böhmer et al. (Böhmer et al., 2015). For each modality as well the additionally identified information is listed and combined and duplicates are removed. This resulted in a list of different data that is structured and layered by designing the common data model.

The first step taken is defining the entities and assigning the related attributes to the entities. The list of entities is shown in table 19. Based on the creation of relationships between the entities the common data model is designed. The common data model can be divided into three parts based on the identified data types in section 5.2. The static order data includes the shipment related information and is mainly retrieved from back-office systems. It includes the information necessary for the automatically retrieval of information from websites. The semi real-time data includes information related to a leg. A shipment has one or more legs. A leg is assigned to one modality and vehicle that does the transportation. The information is mainly retrieved from carrier or terminal websites or is updated based on the information retrieved from a trip level. The real-time data is related to a trip and includes the information as the current geographical position and the real-time data retrieved from disruption sources.

The current structure of the data model ensures that shipments and legs with different modalities can be threatened in an unambiguously way resulting that information showed to the user is the same and in the same format for all modalities. The result of the design process is shown in figure 29 and will be validated in the next chapter.

Chapter 6 Validation of the Common Data Model

In this chapter, the Common Data Model as designed in Chapter 5 will be validated and will, combined with chapter 5, the fifth sub-question. The chapter belongs to the second part of the ADR. Section 6.1 introduces the chapter. The designed CDM has to be validated to check whether it can process data from different sources and in different formats. Section 6.2 the different ways of validation are introduced while section 6.3 discussed the results. Section 6.4 is the concluding section of this chapter.

6.1 Introduction

In chapter 5, the Common Data Model is designed based on the processes and needs of planners at Neele-Vat, HST Groep and CTT. However, to check whether the model is usable for more than just these three organizations and whether the model fits the information that is needed from all the websites, the model has to be validated. Validation is checking that the designed model meets specifications and fulfills its intended purpose. There are multiple methods to validate the CDM since the CDM has to be applicable with messages from different sources. These methods are introduced in section 6.2 while in section 6.3 the CDM is validated with the introduced methods.

6.2 Validation Methods

This section elaborates about different approaches to validate the CDM. It will validate not only the model but also whether the CDM satisfies the requirement of section 5.3. It will check that the CDM holds all the information can be retrieved and required by the logistics providers. These approaches are outcomes of interviews with employees at Neele-Vat, HST, CTT and CAPE Groep, observations at Neele-Vat, HST and CTT and the literature review.

First, the bottom-up approach is discussed. This approach takes different definitions at HST, Neele-Vat and CTT into account to validate the CDM by checking whether all information needed for track and tracing fits in the CDM. The second approach is the industry standards approach. This approach starts with the different logistics standards and checks whether it fits with the CDM and the static dossier data can be derived from these standards. The third approach is the website approach. This approach takes the different data sources into account and validates it with the CDM. The last approach is the CAPE Projects approach. This approach takes relating projects within CAPE Groep to check whether the CDM fits within these projects.

6.2.1. Bottom-Up Approach

The approach shows the applicability of the CDM with the order definition messages of the collaborating companies. This approach starts with different definitions of orders of different companies and messages exchanged of all the modalities. These messages and definitions are received via the HST Groep, Neele-Vat and CTT. Via the processes and interviews described in chapter 3 the data within these messages and definitions that are related and necessary for track and tracing are identified. These definitions are then combined by relating them and find any overlap and compared with the CDM.

6.2.2. Industry Standards Approach

The second approach researches the applicability of the CDM with the Industry Standard order definitions. The approach starts with identifying the different standards used in the logistics industry. There are tens to hundreds of different standards to define orders defined by different organizations. A sub-set of these standards is mapped to the CDM to check whether the standards covers the data necessary to start the information retrieval.

6.2.3. Website Approach

The third approach shows whether the data that is required to retrieve from the World Wide Web can be retrieved. This approach focuses on the data available on the internet. On the internet websites of carriers, terminals, airports etc. information is shown about their processes and statuses of goods and deliveries. During this approach the starting point is the websites that the companies use during their tracking processes. For these websites the data needed will be identified. The websites analyzed are the 14 websites that are analyzed as well in chapter 3 plus an additional ADS website. In table 20 the websites are shown. For each website the data that are necessary to retrieve shipment information is checked with the CDM and whether the retrieved shipment information can be stored in the CDM.

Table 20: Websites analyzed for the website approach

Website Type	Website	Website Type	Website
Carrier	www.maersk.com (Maersk)		http://ecom.kline.com (K-Line)
	www.afklcargo.com (KLM/Air France)		http://www2.nykline.com/ (Nyk Line)
	www.qrcargo.com (Qatar Airways)	**Terminal**	eservices.ect.nl/ (ECT Terminal)
	www.yml.com.tw (Yang-Ming)		www.apmtrotterdam.nl (APMT 1 Terminal)
	tang.csair.com (China Southern Airlines)	**AIS**	www.marinetraffic.com
	www.turkishcargo.com.tr (Turkish Airlines)	**ADS**	www.flightaware.com
	cargo.virgin-atlantic.com (Virgin Atlantic Airways)		www.flightradar24.com
			www.flightview.com

6.2.4. CAPE Projects Approach

The last approach focuses on a related project done by CAPE Groep, the University of Twente and Seacon. This project had similar features and intentions that are more elaborated in this thesis. The project included a fuel-based Logistics Carbon Management System that monitors and collects real-time data about the fuel consumption during trips, and, consequently, calculates detailed and accurate carbon footprints of transportation services. Iacob & van Sinderen et al. (M. E. Iacob et al., 2013) proposed an architecture for this system with a relating data model. The data model will be used by this approach to map this project with the designed CDM.

6.3 Validation

In this section, the CDM will be validated using the four methods explained in section 6.2. The first method will be the bottom-up approach. During the different approaches the CDM is mapped to the different definitions, websites or approaches. For each entity will be given if the attribute of the CDM can be mapped to the definition. If so, the mapping will be shown in related cell of the column 'Source', otherwise the cell will be empty. This means that in the definition, website or project this specific attribute of the CDM is not found.

6.3.1. Bottom-up Approach

To validate the CDM with the bottom-up approach order definitions are needed from HST Groep, Neele-Vat and CTT. These order definitions of the three companies are tested for their applicability with the designed CDM. Important for the order definitions is that the static order data (with some exceptions as ship name) can be derived from the order definitions to create a basic order file in the application. This 'dossier' data are used to retrieve information from websites. This means that for the bottom-up approach it is important that an order can be defined and that the information that is needed as input information for the website calls is in the definition. Therefore, only the static data are mapped and presented in this approach. Each company will be explained separately, starting with HST Groep.

HST Groep

HST Groep defines an order in the combination of several XSD files. Each of these XSD files are examined to map the attributes to the attributes in the CDM. In Table 21 the mapping of HST is presented. What stands out is that the order definition stores no information about the owner/creator of the shipment. This means that from an incoming order the owner of the shipment can't be identified. Another thing is that the order definition does not contain the container number. This means that the container number must be derived via the B/L information, when this is known. Last point is that the modality is not known in the order definition. This can only be identified based on the B/L or AWB Number. For the rest of the data needed at the CDM the order definition can provide that.

Neele-Vat

In comparison with HST Groep, Neele-Vat defines an order in one order message. The order is defined in a sub-set of an IFCSUM EDIFACT message. In the Industry Standards approach more about IFCSUM Messages. In table 22 the contents of the order message is mapped to the CDM. It stands out that all the information is present to define the order and start track and tracing. If Neele-Vat identifies the departments and employees in the application, then via the employee and department unique ID's the department and employee can be linked to the right orders. (Only some ID's or names are missing, but there is always another name or ID present that can fill the missing information.)

Table 21: Mapping of the order definition of HST Groep

Static Data		
Entity	**Attribute**	**Source**
Company	CompanyID	-
	CompanyName	-
Department	DepartmentID	-
	DepartmentName	-
Employee	EmployeeID	-
	EmployeeName	-
Shipment	ShipmentID	-
	OrderType	transportOrderType
	From	TransportOrder - FromAddress - City
	To	TransportOrder - ToAddress - City
	ETD	TransportOrder - ShippingDate
	ETA	TransportOrder - IndicatedDeliveryDate
	Containernumber	-
Customer	CustomerID	-
	CustomerName	TransportOrder - Customer
Leg	Modality	-
	From	TransportOrder - FromAddress - City
	To	TransportOrder - ToAddress - City
	ETD	-
	ETA	-
	BL_Number	TransportOrder - BillOfLadingNumber
	AWB_Number	TransportOrder - BillOfLadingNumber
Carrier	CarrierID	-
	CarrierName	TransportOrder - Subcontractor/type

Table 22: Mapping of the order definition of Neele-Vat

		Static Data
Entity	**Attribute**	**Source**
Company	CompanyID	-
	CompanyName	-
Department	DepartmentID	-
	DepartmentName	OrderGemaaktDoorAfdeling
Employee	EmployeeID	-
	EmployeeName	OrderGemaaktDoor
Shipment	ShipmentID	OrderNummer
	OrderType	OrderType
	From	SoortTransport
	To	PlaatsVanHerkomst
	ETD	ETD
	ETA	ETA
	Containernumber	ContainerNummer
Customer	CustomerID	KlantReferentienummer
	CustomerName	-
Leg	Modality	SoortTransportType
	From	SoortTransportHerkomst
	To	SoortTransportBestemming
	ETD	ETD
	ETA	ETA
	BL_Number	DouaneZeeBL
	AWB_Number	DouaneLuchtvrachtBL
Carrier	CarrierID	-
	CarrierName	Vervoerder/RederijZoeknaam

CTT

From CTT an order definition is received. This order is defined in a sub-set of an IFTMIN EDIFACT message. Several experts on EDIFACT messages helped during the mapping of the definition but unfortunately, no mapping could be realized. The definition uses too much own defined fields that could not be identified using the IFTMIN specifications. No experts within CTT were available to help identifying the fields. This means that no mapping is made for CTT.

6.3.2. Industry Standards Approach

For industry standards approach it is important to identify the information in the Standard related to the information stored in the CDM. Just as in the Bottom-up Approach the static order definition is important since the Standards used are order definitions. For validation three Industry Standards are tested, IFCSUM, IFTMIN and GS1 Standard Transport Instruction. All messages are an EDIFACT messages. EDIFACT is an international standard to exchange messages electronically and is defined by the United Nations. It is used by a lot of transportation companies. Firstly, the IFCSUM messages explained.

IFCSUM

An IFCSUM message is a summary message used in Electronic Data Interchange (EDI) between trading partners involved in administration, commerce and transport. It is used by parties to define and exchange order messages. Table 23 shows a mapping of an IFCSUM message with the CDM. A note to the mapping is that this might be not complete since transportation companies can use different codes to mean the same thing. Identifying all possibilities is not possible because transportation companies misuse a lot of field but table 23 shows a possibility for a mapping. Nothing really pops out when mapped, everything can be mapped into the CDM. For (almost) every field is a field in the IFCSUM. Only for some ID's it is not possible to map, but since the names are mapped, the ID can be retrieved via the name.

Table 23: Mapping of the order definition of EDIFACT IFCSUM

Static Data

Entity	**Attribute**	**Source**
Company	CompanyID	-
	CompanyName	-
Department	DepartmentID	SG15 CTA C056 3413 Department or employee identification DE/DI + Other codes
	DepartmentName	SG15 CTA C056 3412 Department or employee based on ID code
Employee	EmployeeID	SG15 CTA C056 3413 Department or employee identification DE/DI/IC + Other codes
	EmployeeName	SG15 CTA C056 3412 Department or employee based on ID code
Shipment	ShipmentID	SG1 RFF C506 1154 Reference number AAS
	OrderType	SG15 CTA C056 3413 Department or employee identification DE/DI (only Import/Export)
	From	SG13 NAD 3164 City name + Country CZ / SF
	To	SG13 NAD 3164 City name Country CN / ST
	ETD	DTM C507 2380 Date/time/period Codelist 2005 2 / 17 Delivery Date/time estimated Last Trip
	ETA	DTM C507 2380 Date/time/period Codelist 2005 200/196
	Containernumber	SG12 EQD C237 8260 Equipment identification number
Customer	CustomerID	-
	CustomerName	SG13 NAD C080 3036 Party name
Leg	Modality	SG1 RFF C506 1153 Reference qualifier MB
	From	SG13 NAD 3164 City name Departure Trip
	To	SG13 NAD 3164 City name Arrival Trip
	ETD	-
	ETA	-
	BL_Number	SG1 RFF C506 1154 Reference number based on 1153 Reference qualifier MB
	AWB_Number	SG1 RFF C506 1154 Reference number based on 1153 Reference qualifier AWB
Carrier	CarrierID	-
	CarrierName	SG13 NAD C080 3036 Party name

IFTMIN

Both IFCSUM and IFTMIN are EDIFACT messages but where IFCSUM a summary message is IFTMIN is the definition of an Instruction Message. It is suitable for the arrangement of the transport between all parties. Part of the message consists of how the order defined. Table 24 shows the mapping of an IFTMIN message. As with the IFCSUM message, parties misuse often fields which means this mapping is not complete, but only shows how the messages should be filled according to the official information. IFTMIN can contain all the information necessary to retrieve order data. As with IFCSUM not all ID's are filled out, but since the names are mapped, the ID can be retrieved via the name in the application.

Table 24: Mapping of the order definition of EDIFACT IFTMIN

	Static Data	
Entity	**Attribute**	**Source**
Company	CompanyID	-
	CompanyName	-
Department	DepartmentID	CTA C056 3413 Department or employee identification DE/DI + Other codes
	DepartmentName	CTA C056 3412 Department or employee based on ID code
Employee	EmployeeID	CTA C056 3413 Department or employee identification DE/DI/IC + Other codes
	EmployeeName	CTA C056 3412 Department or employee based on ID code
Shipment	ShipmentID	SG3 RFF C506 1154 Reference number AAS
	OrderType	CTA C056 3413 Department or employee identification DE/DI (only Import/Export)
	From	SG13 NAD 3164 City name + Country CZ / SF
	To	SG13 NAD 3164 City name Country CN / ST
	ETD	DTM C507 2380 Date/time/period Codelist 2005 2 / 17 Delivery Date/time estimated Last Trip
	ETA	-
	Containernumber	SG37 EQD C237 8260 Equipment identification number
Customer	CustomerID	-
	CustomerName	SG11 NAD C080 3036 Party name
Leg	Modality	SG3 RFF C506 1153 Reference qualifier MB
	From	SG11 NAD 3164 City name Departure Trip
	To	SG11 NAD 3164 City name Arrival Trip
	ETD	DTM C507 2380 Date/time/period Codelist 2005 10 /200/196
	ETA	DTM C507 2380 Date/time/period Codelist 2005 17 Delivery Date/time estimated Trip / 2
	BL_Number	SG3 RFF C506 1154 Reference number based on 1153 Reference qualifier MB
	AWB_Number	SG3 RFF C506 1154 Reference number based on 1153 Reference qualifier AWB
Carrier	CarrierID	-
	CarrierName	SG11 NAD C080 3036 Party name CA

GS1

The last standard is the GS1 Standard Transport Instruction defined by the Dutch GS1 organization. This standard is a subset of the IFTMIN message, and is a popular standard for Dutch transportation companies. Table 25 shows the mapping with the GS1 Standard Transport Instruction message. Most of the information in this standard is present to define the order, although it is not possible to relate the order its owner. Other important information that is missing is that the modality is not given for an order. This should be retrieved via the AWB or B/L number.

Table 25: Mapping of the order definition of GS1 Standard Transport Instruction

	Static Data	
Entity	**Attribute**	**Source**
Company	CompanyID	-
	CompanyName	-
Department	DepartmentID	-
	DepartmentName	-
Employee	EmployeeID	-
	EmployeeName	-
Shipment	ShipmentID	Transportordernummer = Ordernummer
	OrderType	-
	From	Haaladres Eerste Leg-Plaats
	To	Afleveradres Laatste Leg -Plaats
	ETD	Gewenste leverdatum/tijdstip
	ETA	Gewenste afhaaldatum/tijdstip
	Containernumber	Equipmentidentificatie
Customer	CustomerID	Factuuradres - GLN
	CustomerName	Factuuradres - Naam
Leg	Modality	-
	From	Haaladres-Plaats
	To	Afleveradres-Plaats
	ETD	-
	ETA	-
	BL_Number	Zendingnummer
	AWB_Number	Zendingnummer
Carrier	CarrierID	Vervoerder-GLN
	CarrierName	-

6.3.3. Website Validation

The third way to validate the CDM is the website validation. Modalities Deep Sea and Barge are taken together while Air is taken separately. For every modality is tested whether the information we need from a type of website, for example Carrier websites or Terminal websites, to update a shipment can be placed into the CDM.

For the websites the semi real-time data and real-time data are important, expect for the disruption and status message because these messages are created based on the data retrieved from the websites. These messages can send messages back to the back office systems and are therefore not important here. Therefore only the semi real-time and real-time data are mapped and shown in the tables.

Deep Sea and Barge

For the websites for modalities Sea and Barge three types of websites are used. Firstly, the website of the carrier where the schedule of a ship can be retrieved as well the B/L information of the shipment. Secondly, the Terminal websites from Rotterdam that have status updates about the arrival of a ship. At last, the Marinetraffic website that stores AIS data. Table 26 shows that most information can be derived from the websites. Only for the weather information is not all the weather attributes can be derived.

Table 26: Mapping of websites for sea and barge

Semi-Real Time Data			Real-Time Data		
Entity	**Attribute**	**Source**	**Entity**	**Attribute**	**Source**
Shipment	Status	Based on Legs	**Location**	Latitude	Marinetraffic
Leg	ETD	Carrier website		Longitude	Marinetraffic
	ETA	Carrier website		Course	Marinetraffic
	Status	Carrier website/ Terminal website		Date_Timestamp	Marinetraffic
	ContainerStatus	Carrier website/ Terminal website	**Weather**	Temperature	Marinetraffic
				Windspeed	Marinetraffic
Vehicle	VehicleID	Carrier website		Winddirection	Marinetraffic
Ship	ShipName	Carrier website		WeatherType	-
Airplane	Flightnumber			Humidity	-
Trip	From	Carrier website		Visibility	-
	To	Carrier website		Location	-
	ETA	Carrier website			
	ETD	Carrier website			
	Status	-			

Air

For the Air modality two types of websites are used for retrieving information. Firstly, the website of the air carrier. Here information is shown about the schedule and the status of the container. The second type of website is Flightradar that is used to follow current flights and retrieve current information. A third website is used (Flightview) to retrieve the weather information at the different airports. The mapping is shows in Table 27. It shows from what websites different data can be retrieved. Looking again to the second and third set of columns, it can be concluded that all information necessary can be retrieved.

6.3.4. Cape Projects Validation

During the Cape Projects validation the SEACON project will be used as validation to check whether this projects has the information needed for the track and tracing application. If it does, it might be possible to integrate the projects. The designed CDM will be validated by mapping the attributes to attributes to the domain model of the SEACON project of Iacob & van Sinderen et al. (M. E. Iacob et al., 2013).

The mapping is shown in tables 28 and 29. The conclusion from this mapping is that this projects stores not the right information to directly track and trace the shipments stored in the SEACON Product Carbon Footprint Calculator domain model. Essential information is missing as OrderType, Container number, BL- and AWB Number and ETA and ETD information at Leg-level. This is critical information that is needed for the application to work properly.

Table 27: Mapping of websites for air

	Semi-Real Time Data			**Real-Time Data**	
Entity	**Attribute**	**Source**	**Entity**	**Attribute**	**Source**
Shipment	Status	Based on Legs	**Location**	Latitude	Flightradar24
Leg	ETD	Carrier website		Longitude	Flightradar24
	ETA	Carrier website		Course	Flightradar24
	Status	Carrier website		Date_Timestamp	-
	ContainerStatus	Carrier website	**Weather**	Temperature	Flightview
				Windspeed	Flightview
Vehicle	VehicleID	Carrier website		Winddirection	Flightview
Ship	ShipName			WeatherType	Flightview
Airplane	Flightnumber	Carrier website		Humidity	Flightview
Trip	From	Carrier website		Visibility	Flightview
	To	Carrier website		Location	Flightview
	ETA	Carrier website			
	ETD	Carrier website			
	Status	-			

Table 28: Static data mapping for SEACON project

Static Data					
Entity	**Attribute**	**Source**	**Entity**	**Attribute**	**Source**
Company	CompanyID	-	**Customer**	CustomerID	-
	CompanyName	-		CustomerName	-
Department	DepartmentID	-	**Leg**	Modality	Trip-Modality
	DepartmentName	-		From	PickupLocation-LocationID
Employee	EmployeeID	-		To	DeliveryLocation-LocationID
	EmployeeName	-		BL_Number	-
Shipment	ShipmentID	Shipment-ShipmentID		AWB_Number	-
	OrderType		**Carrier**	CarrierID	Shipper-ShipperID
	From	PickupLocation-LocationID		CarrierName	Shipper-ShipperName
	To	DeliveryLocation-LocationID			
	ETA	Shipment-DeliveryDate			
	ETD	-			
	Containernumber	-			

Table 29: Semi real-time and real-time data mapping for SEACON project

Semi Real-Time Data			Real-Time Data		
Entity	**Attribute**	**Source**	**Entity**	**Attribute**	**Source**
Shipment	Status	Shipment-Status	Location	Latitude	-
Leg	ETD	-		Longitude	-
	ETA	-		Course	-
	Status	Trip-Status		Date_Timestamp	-
	ContainerStatus	-	Weather	Temperature	-
				Windspeed	-
Vehicle	VehicleID	Vehicle-VehicleID		Winddirection	-
Ship	ShipName	-		WeatherType	-
	ShipStatus	-		Humidity	-
Airplane	Flightnumber	-		Visibility	-
Trip	From	-		Location	-
	To	-			
	ETA	-			
	ETD	-			
	Status	-			

6.4 Conclusion

Chapter 6 validated the in chapter 5 designed CDM and answered completes the sixth sub-question:

> *SQ6: How is a common data model for planned and actual information, orders and status/disruptions designed?*

The CDM and its requirements are validated via four different methods: Bottom-up, Industry Standard, Website Validation Approach and CAPE Projects Approach. The Bottom-up Approach validates the CDM by testing the order definitions of the three companies for their applicability with the designed CDM. For HST Groep, their order definitions contains the basic information that is needed to track and trace the shipments, but the shipment cannot be assigned to an employee. Neele-VAT has an order definition that includes all information necessary for the application. For CTT, it was not possible to make a mapping.

In the Industry Standards Approach the definition of various definitions are compared to the CDM. IFCSUM and IFTMIN messages can be mapped on the CDM, although it is important to watch out that companies won't use different field to store the same information. Since both messages are very large it is possible to use different fields for the same information. GS1, a subset of the IFTMIN message, is a message that cannot be mapped fully to the CDM and therefore not less useful since it misses important information as the modality of the shipment.

The third approach is the Website Validation Approach. During this approach it is tested for every modality if the information that the dossier needs is available on the related websites. For Sea, Barge and Air the information needed is available. Only for Sea and Barge some weather attributes cannot be filled.

The last approach is the CAPE projects approach. In this approach the CDM is mapped to the data model of a project at CAPE which is mentioned in the paper of Iacob et al. (M. E. Iacob et al., 2013). With the current model the information stored in the designed CDM cannot be stored. Essential information is missing and in order to work the data model of the SEACON project must be changed.

An overview of the applicability of the different definitions, websites and projects is shown in table 30.

Table 30: Applicability of the different definition, websites and projects

Approach	Name	Applicability
Bottom-up Approach	HST	+
	Neele-Vat	++
	CTT	--
Industry Standards Approach	IFTMIN	++
	IFCSUM	++
	GS1 NL	+/-
Website Approach	Websites	+
CAPE Projects Approach	SEACON	-

Chapter 7 Prototyping the CDM into an Integration Platform

This chapter describes the design and development of an Integration Platform which uses the designed CDM, described in Chapter 5, and architecture of Chapter 4. This chapter is the second last chapter of the second step of the ADR and will answer the sixth sub-question. The Integration Platform will be used to test the CDM and the functionalities of the platform in practice. It will track and trace shipments automatically. The platform will be a prototype to show the advantages of the CDM and platform in comparison with the current situation. In section 7.1 the chapter will be introduced. Section 7.2 introduces the architecture for this prototype, based on the architecture from Chapter 4. In sections 7.3 – 7.5 the design and design choices of the different applications or services are discussed. Section 7.6 described the overall performance of the prototype. The prototype is evaluated and validated in section 7.7. The chapter ends with a conclusion in section 7.8.

7.1 Introduction

In the previous chapters, several transportation companies are visited and the current situation at these companies are observed. Interviews with employees are held to collect information about the data they use, how decisions are made and how their track and trace processes look like. These interviews PE resulted in a conceptual design and a list of functionalities for the prototype. The conceptual design is shown in figure 33.

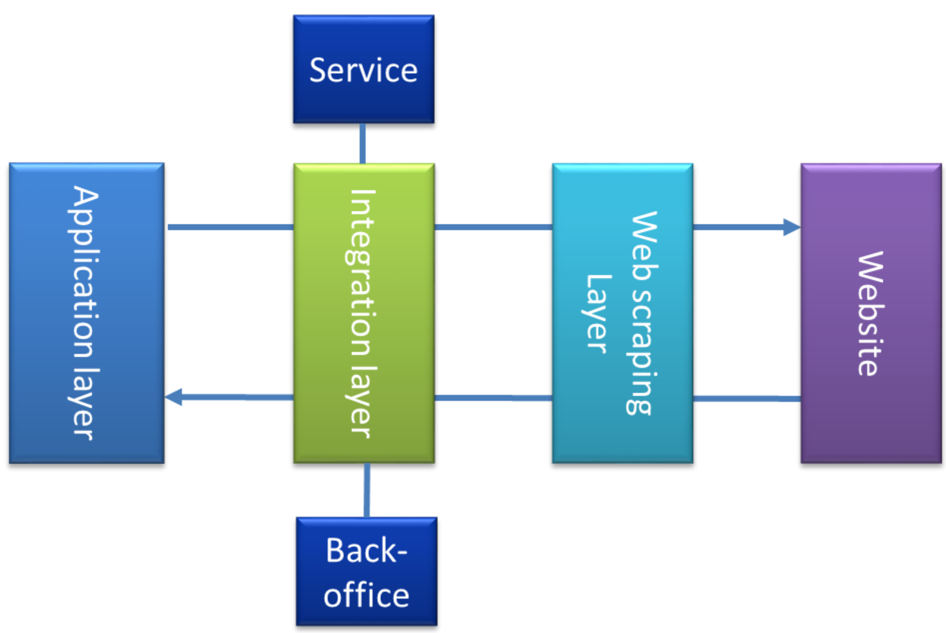

Figure 33: Conceptual design of the integration platform

In chapter 4 the target architecture is designed including the integration platform that consists of multiple applications and services. In the application layer, a track and trace service is modeled that takes care of the user interface and the communication with the inputs of the user to the rest of the platform. This application is the application layer in the conceptual design.

The integration bus is modeled in the technology layer and has as function to route and transform the messages. Every system is connected to the bus which makes the bus the integration layer. Two other services are added to the conceptual design: Service and Back-office. Both are not in the scope of the prototype but are included in the architecture since they are needed for the functionalities of the platform. Back-office integration is necessary to run the application independently without inputs from the end-user. These back-office systems feed the platform and application with the order data. The platform has a connection with service as well. It enables to link customers of transportations companies to the platform so they can follow the status of their own shipments. Customers are interested in the progress of their shipment and this feature enables them to have visible information at each moment in time without contact the transportation company.

Cloudscrape is the web scraping tool used in this research and makes sure that the data is scraped from the websites. This is the web scraping layer. In the application layer a Mendix application (see section 7.3) will be built that process the incoming data and is the application used by end-users. For the integration layer eMagiz is used (see section 7.4) to create an integration bus. The web scraping tool used is Cloudscrape (see section 7.5). Each part has its own design process and choices and will be elaborated in different chapters. During the chapter, each of these applications/services will be more elaborated. In the following sections, the architecture of the prototype and the three parts of the prototype (Mendix, eMagiz and Cloudscrape) will be elaborated, starting with the architecture of the prototype. In the other sections, the tools will be introduced, the relationship with the research, the translation to the prototype and the issues will be discussed.

7.2. Architecture of the prototype

In Chapter 4 the architecture is designed for after the implementation of the platform. The prototype will not be a fully complete platform. Some details are left out and for the prototype is chosen to focus on the core functionalities to prove the concept of automatically update shipments using web scraping tools and data. The prototype will include the basic functionalities that are necessary to show planners the usage and the advantages of the platform and enable planners to use the platform to track shipments within the constraints of the added websites.

Figure 34 shows the architecture of the prototype as it is built. It has many similarities with the future architecture in chapter 4 although some websites, functionalities and processes are left out because of the earlier mentioned reason. This means for example that connections to customers are left out, but can be easily added.

In the architecture, the Application layer represents the Mendix application and in the Technology layer, the eMagiz bus represents eMagiz and Cloudscrape API represents Cloudscrape. The different functions and application components of Mendix will be discussed in the next section.

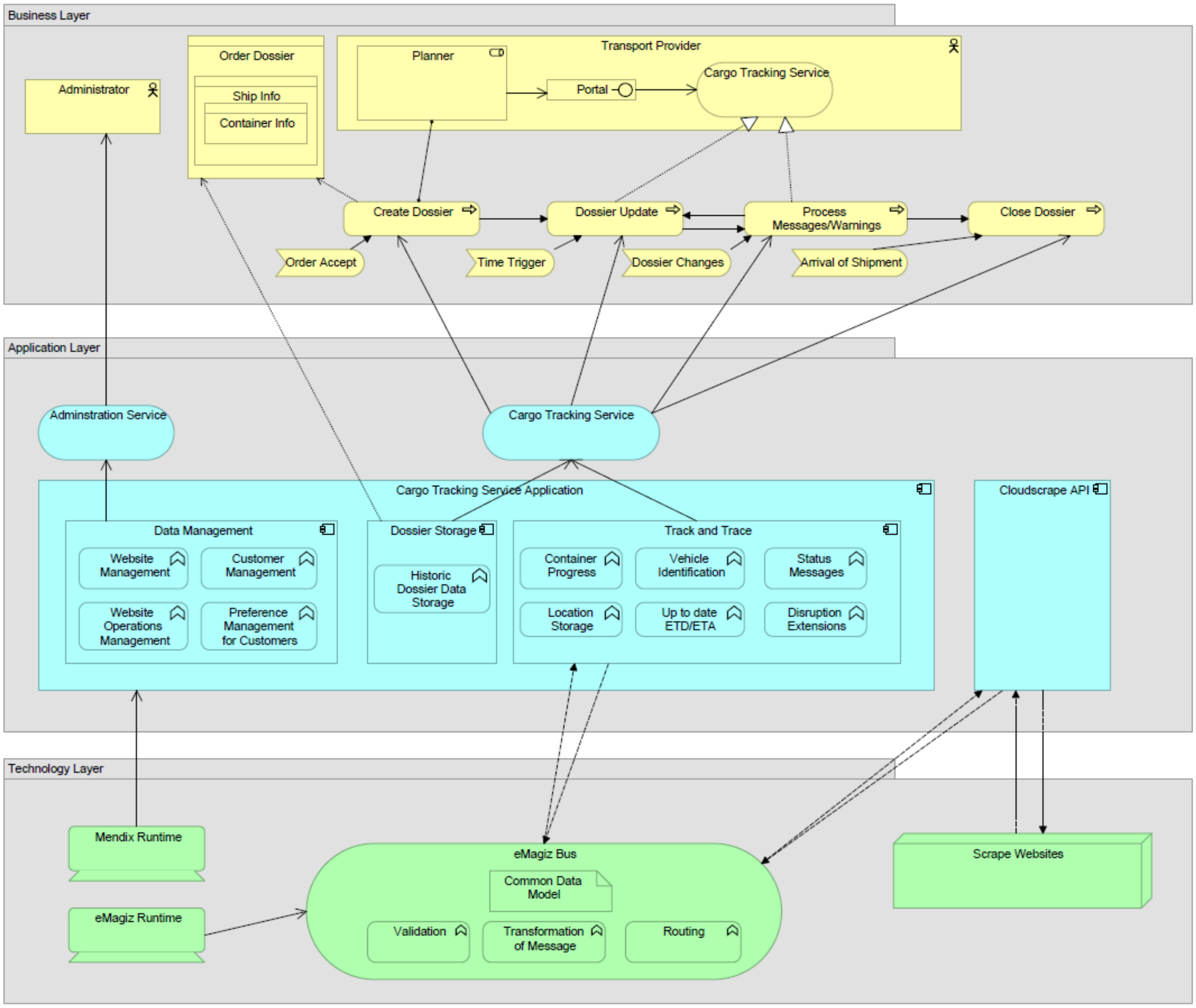

Figure 34: The architecture of the prototype

7.3. Mendix Application

In this section, the Mendix application, part of the prototype, will be discussed. The Mendix application is situated in the Application Layer in the architecture.

7.3.1. Mendix Introduction

Mendix is an application Platform as a Service (aPaaS), which can be applied to develop applications quickly, without using and knowledge coding languages. Withing CAPE Groep the use of Mendix is common and, therefore, the choice to use Mendix as platform is a logical one. It offers the possibility to build an application based on models and process flows while the application can be deployed easily and quickly for testing purposes.

7.3.2. Connection to Research

In Chapter 1 is defined that the research goal to improve the ability to execute synchromodal modality shifts by designing a common data model for planned and actual information to structure unstructured and structured data in order to make data sharing among the supply chain easier.

An application is needed for the planners of the transportation companies where information is stored and presented to make more efficient and faster modality shifts and decisions. In this application, the unstructured data from websites are structured and presented in a generic way for every modality to the user. The application is also used to store planned, actual and historical information connected to a shipment and will update the shipments automatically on a more regular base than currently happens manually. This enables faster recognition of disruptions and status changes and enables more efficient modality shifts. Customers of the transportation companies can be connected to their own personal page here they can automatically be informed about the status and progress of their shipments.

7.3.3 Application Functionalities

This section will elaborate about the functionalities of the prototype. Figure 35 shows an overview of the required functionalities that will build for the prototype. During several conversations with CAPE Groep, it became clear that the main focus next to the regular track and tracing of containers should be the flexible management system of companies, websites and their operations. This management system should be able to assign companies to the websites they prefer to use for their shipments. Every company has contracts with different carriers and they don't prefer to be connected to all of the different carriers. Another argument for this flexible system is that websites can be updated and that problems must be solved without many proceedings. This can cause problems at the designed web scrape robot. For example, it may be that the websites require for the same operation different information from the prototype. The prototype should also be scalable. Therefore, it requires that new websites and operations are easily added to the database and can be assigned to customers.

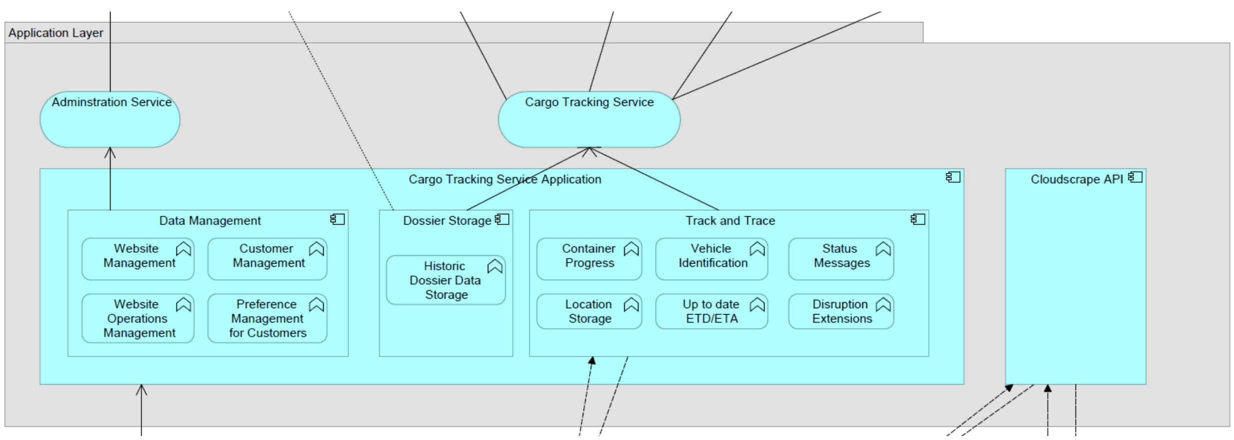

Figure 35: The application overview of the prototype

7.3.3.1. Track & Trace Functionalities

The application functionalities are at the heart of the prototype. Their part is updating the dossiers by requesting data from the web scrape robots and process this by business logic. The functionalities described are partly business logic as design choices for scalability of the prototype. The functionalities are:

- <u>Progress Container</u>
  The prototype will calculate the progress of the container in percentages. Based on this percentage it will be determined what websites are preferred for certain data as ETA or ETD. This

percentage will also give a good overview to the end user who will interpret the progress of their shipment easily.

- Location Storage
  The prototype is able to retrieve coordinates of the position of the vehicle that transports the containers. The location coordinates will be stored and can be used for data mining purposes when a requirement for analyzing data comes into the scope of the prototype. The coordinates are also used to show the current position of the vehicle.
- Vehicle Identification
  A very important function is the vehicle identification function. Carriers switch containers to different ships without acknowledging their customer. This can have a big difference in the arrival time of the container, the route of the ship and its arrival location. Notifying this change of vehicle is important since others processes might be started by the end user. For example, the end user must arrange transportation from another port or have to notify another transportation company who picks up the container at the port that the container arrives later or earlier.
- Up to date Status/ETA/ETD
  The Status, ETA and ETD information are for end users the basis to start and end processes. Based on this data they can order the transportation from the arrival location, inform the customer and the hired transportation company.
- Disruption Extensions
  Disruption messages are sent to the owner of the dossier when some data in the dossier is updated exceptional within the set borders. This enables the owner to calculate the risks and take action if necessary. The model is extensible so if decided to extend the data model to store and use more data from new data sources, this is possible. This way, new disruptions can be added, like traffic information for trucks.
- Update Messages
  Next to disruption messages are the regular update messages that inform the owner about the progress of the shipment. These messages do not have the intention to warn the owner but are informal and contain information like status changes or that the ETA of a shipment is within a week of arrival.

7.3.3.2. Dossier Storage Functionalities

Next to the track and trace functionalities are the dossier storage functionalities which are meant to save all the data relating to a shipment in a database. The decision is made to store all this data to build a database that can be used in the future for data mining purposes. However data mining and analyzing are not in the scope, it is a very powerful method to discover additional and important information about the performance of transportation companies, carriers and ships, what routes have bottlenecks etcetera. To be able to perform data analytics, a database must be built with real-time data.

7.3.3.3. Data Management Functionalities

The last group of functions is the functionalities that are related to data management. These functions are important in the backend of the system where the master data are managed and flexibility features are added. In the backend of the system websites and its operations are defined as well as the sequence the data are scraped for each combination of modality and (if necessary) import, export and inland transportation. The transportation companies are linked to the websites of their preference.

Each functionality is further explained in the following functions:

- **Manage Websites**
  This feature covers the management of the websites and the flexibility to sequence the websites for each modality and (if necessary) import, export and inland transportation. For each combination, a different sequence of websites is preferred to minimize the number of robot executions which has a big influence on the performance of the prototype. For example, for deep sea transportation, the import process wants to retrieve information from the terminals in Rotterdam at the end of the leg while the export process from Rotterdam wants to retrieve the information at the beginning of the leg.
- **Manage Website Operations**
  Each website stores all kinds of information and each kind of information can be retrieved by a web scraping robot defined and built in Cloudscrape. An operation defines the type of information retrieved from the website and relates the operation with the Cloudscrape robot. For each website, the sequence of the operations can be defined and changed any time so that a sequentiality is defined to retrieve the information first before process the other operations.
- **Preference Management for Customers**
  One of the important features in the prototype is the preference management feature for customers who are connected to the prototype. Each customer is assigned to its desired websites and modalities and able to change the sequence of the websites and is able to manage the frequency of the update calls within their own set borders. This enables customers to define their own number of updates to their own preferences and experiences.
- **Customer Management**
  The owner of the prototype should be able to manage their customers and the connections they have with selected websites. This feature enables this desire by letting the owner add, delete or change the access to websites by the combination of modality, import/export/inland, and the website.

7.3.4. Input

During the design and the build process of the Mendix application different input information collected earlier in the research is used. The information retrieved from the interviews and the process identification of different modalities and order types have been important. This can be found in Chapter 3 and the elaboration of the interviews in Appendix B. The interviews also identified what information is necessary to retrieve from websites. This is translated earlier in the design of a common data model, see Chapter 5. This model has been important as a starting point for my domain model since it holds all the information necessary that needs to be stored for the dossiers/shipments.

7.3.5. Translation to the prototype

Based on the input I retrieved earlier this thesis important data are collected as well important process steps are identified. This subsection aims to explain what choices are made and translated into the application based on the collected input.

In chapter 3 the processes are identified how shipments are tracked for different modalities and order types. For each modality, the process is different based on what information is available, retrieved and the number of sources. In chapter 5 the common data model is designed and shows that each modality is presented in the same format. In Mendix the process to retrieve the data for each modality is generic. The results of the interviews showed that the identification of the vehicle is very important before the vehicle specific information can be retrieved. It is important that this information is retrieved before other data will be retrieved. If all the requests for the data are sent simultaneously and the cargo is on another vehicle, then will the data retrieved from the other requests not valid since these are based on the vehicle that is still stored in the application. Therefore, Mendix its first request for every modality is to retrieve the current vehicle. When the vehicle is identified, the rest of the requests will be sent and collected. When all the information is received by Mendix, the responses are processed. Starting with the information at a trip level. Based on this information and the responses the leg will be updated. If necessary, the shipment will be updated as well. This process is shown in Figure 36.

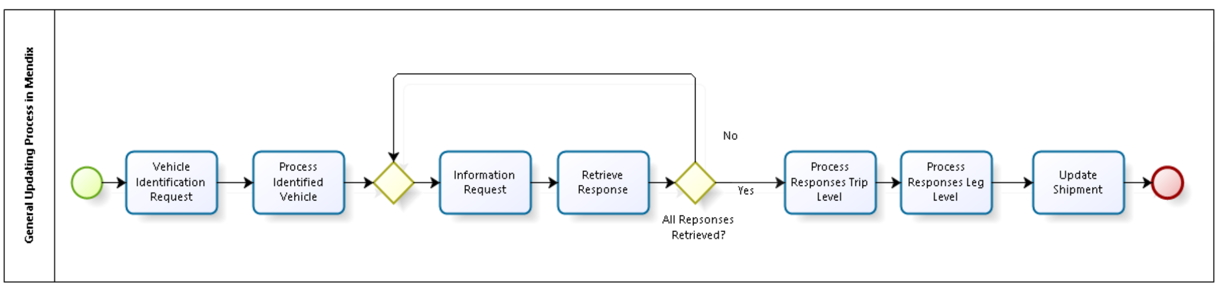

Figure 36: General updating process in Mendix

Mendix identifies the retrieval of the same type of information on different websites, for example, an ETA. According to the interviews the planners choose based on the progress of the leg what source the ETA is optimal or is most reliable. Planners choose currently the source based on the number of days after departure and before arrival. This is translated into the application by adding functionality that based on the progress of the leg the application will automatically choose what ETA is most optimal. This is done for each data type where the same data type is retrieved from multiple websites.

The progress of the leg is a calculated value, based on the ETD, ETA and Current Date and Time. This is done since there are between leg large differences in travel time. For example, deep sea containers can have a travel time of a week but a month is no exception as well. Transportation by air is even faster, that can take hours to days.

Another point of interest identified in the processes are the arrival terminals by deep sea and barge transport. Often the arrival terminal is not determined when the order is planned and accepted which means that the arrival terminal is not known in the dossier of the leg. During the trip, planners are visiting the arrival terminals hoping to identify the arrival terminal the ship will arrive. The application must identify this as well by visiting terminal websites one by one in order to find the terminal of arrival. When found, the application will not search further. The application gives the planner the functionality to determine the sequence the websites are scraped. This means that the planner can speed up the search process by listing the most popular terminals on top.

Planners are currently updating their dossiers on scheduled moments each day, for example, each morning at 9 am. This is the most time-consuming issue since it takes visiting several websites for each shipment in order to update the shipment with the latest information. To minimize the time planners are working with updating the shipments, the application updates the shipments automatically. Planners can see the latest information when they log in into the application and are notified of any warnings. The frequency of updating is not equal to the duration of the shipment. For example, during the departure and arrival of the trip, the frequency is higher than in the middle of the shipment. The application offers the planners to set the frequency of updating for different progress ranges for each modality and order type. This way, the number of requests can be lowered, the performance of the platform will rise while the planner can set the frequency to its own preferences.

The goal of the thesis is to design a common data model for planned and actual information and status/disruptions by enabling to treat different modalities the same way as much as the possible. Since the common data model is re-used as much as possible in the application, it enables to treat the different shipments in a generic way. This is retrieved by designing one overall request and response processing process which is set up generic and is used for all the modalities. Only during the processing of the responses small differentiations are made since vehicles types as ships and airplanes are identified in different notations. In the rest of the application, each modality and each shipment are treated the same way. How this results in generic requests and responses is explained in the section for eMagiz.

7.3.6. Issues

During the development, several issues came up. Most of them have been solved although some are still open. In this subsection, the issues related to the Mendix application will be discussed.

The platform can automatically update legs with modalities Deep Sea, Barge and Air. For testing purposes, a deep sea order is chosen since it has a long travel time, which gave the possibility to test the application on a longer period of time. Negative side effects are that this means that data are updated on websites over a longer time of period, which means that the bug fixing process is really slow and the error handling within the application is not yet completed. It results that the application is not bug-free and still contains some errors.

7.4. eMagiz

This section focuses on the eMagiz bus. In the architecture, it is placed in the Technology Layer as the eMagiz bus as shown in figure 37.

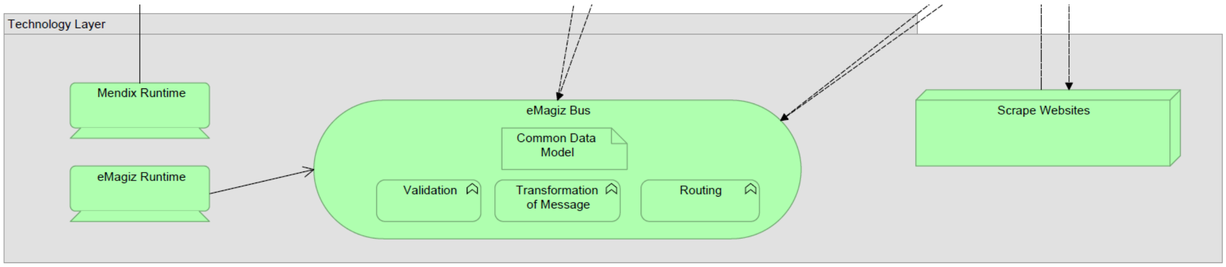

Figure 37: The technology overview of the prototype

7.4.1. eMagiz Introduction

eMagiz is a web-based integration Platform as a Service (iPaas). With it, you can build integration solutions connecting different systems. eMagiz provides functionalities to validate input messages and transform them into an output that receiving systems understand. It integrates with Mendix perfectly as standard modules for this integration are available. eMagiz offers, just as Mendix, users to build model-driven based on flows. Each flow is the translation of a message. The user interface allows the user to easily design and integration systems. It is a perfect tool to connect the Mendix application with Cloudscrape. Another reason to use eMagiz is that it enables easily other systems to the platform, such as back-office transport management systems to the platform.

7.4.2. Connection to the Research

An integration platform is needed to connect Mendix and Cloudscrape. For this connection, a tightly coupled connection could have been chosen but since multiple systems (of other carriers, customers and partners) have to be connected as well, an integration platform is chosen for its scalability and the ability to loosely couple connection between the different end systems. A common data model is central to the process of integrations. In Chapter 2 is defined that ‘in order to integrate different systems with each other their own local data model, the local data schemas have to be transformed into common schemas that are expressed in a CDM’. The Mendix application needs website information to be functional. Information is retrieved via Cloudscrape which means that an integration tool is needed to send requests and responses from Mendix to Cloudscrape and back. The common data model, designed in Chapter 5, is applied in the integration tool and is the data model where the messages are transformed into before transforming to the output format.

7.4.3. Input

The designed common data model is an important input for designing and building the bus. It is used to transform an input message to an output message and to validate if the transformation was successful. Another important input is the generic goal of the thesis. This means that the bus and its processes are designed as generic as possible.

7.4.4. Translation to the Prototype

The bus is designed with two flows. The first flow transports the response to Cloudscrape and the second flows transport the response back to Mendix. An initial design implemented synchronous flows, but since the response from Cloudscrape is not send back immediately when the request arrives one component of the bus runs against a time out. This time-out cannot be changed, so an asynchronous design is chosen.

The flows are designed to be as generic as possible. Research by Veldhuis (Veldhuis, 2015) showed flows for each connected website while the response is not specifically mapped to one request type. Veldhuis put all the data necessary for the request for one request and has a response that includes all possible types of information. Figure 38 and 39 shows the difference between the design of the bus.

Figure 38: eMagiz bus by Veldhuis (Veldhuis, 2015)

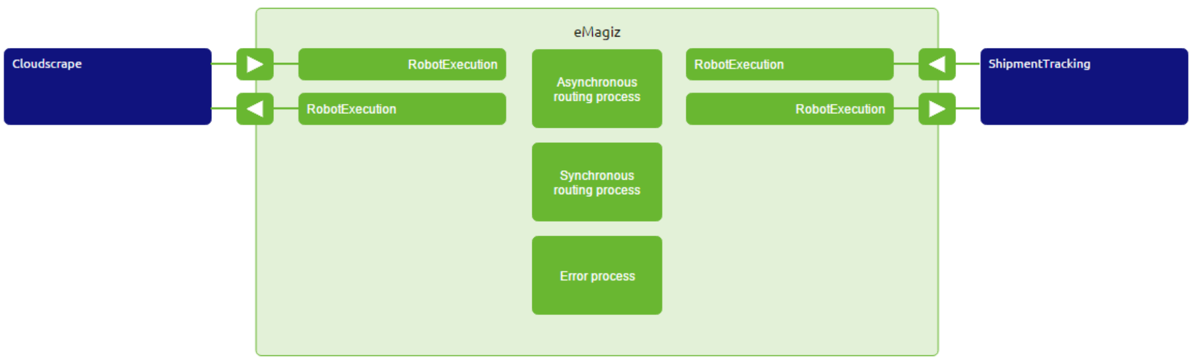

Figure 39: Designed eMagiz bus for the prototype

For this thesis, every request has the same format. As shown in figure 40. It sends only the values that are necessary for the specific robot call. It does not matter what type of request is sent or even when a request is sent from another context, as long as it keeps the format it will be transported to Cloudscrape. The request will be sent to the Cloudscrape API which will initiate a run of the specific robot.

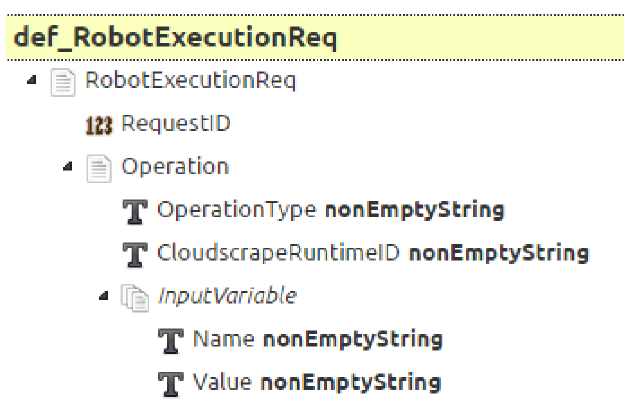

Figure 40: Generic request to Cloudscrape

The response flow is less generic than the request because of the response has to be transformed back to a response that the Mendix application can understand. In the response flow will be dealt with the data conflicts mentioned by Zhang (Zhang, 1994). During the transformations, the format retrieved from the websites is transformed into the common format. Mendix expects different types of responses based on what type of information the application wants to retrieve. Since different requests can have different information in their response, the response is differentiated per request type.

Between the retrieve and sending the request/response to Cloudscrape or Mendix the message is transformed to the (parts) of the CDM meaning that the message is transformed to the CDM and from the CDM back to the format the receiving system supports. Transformations are done by mapping value fields within message format to the CDM and back. The CDM for the responses is shown in Figure 41. It shows types of information that are expected for each request type. The information that is expected is based on CDM.

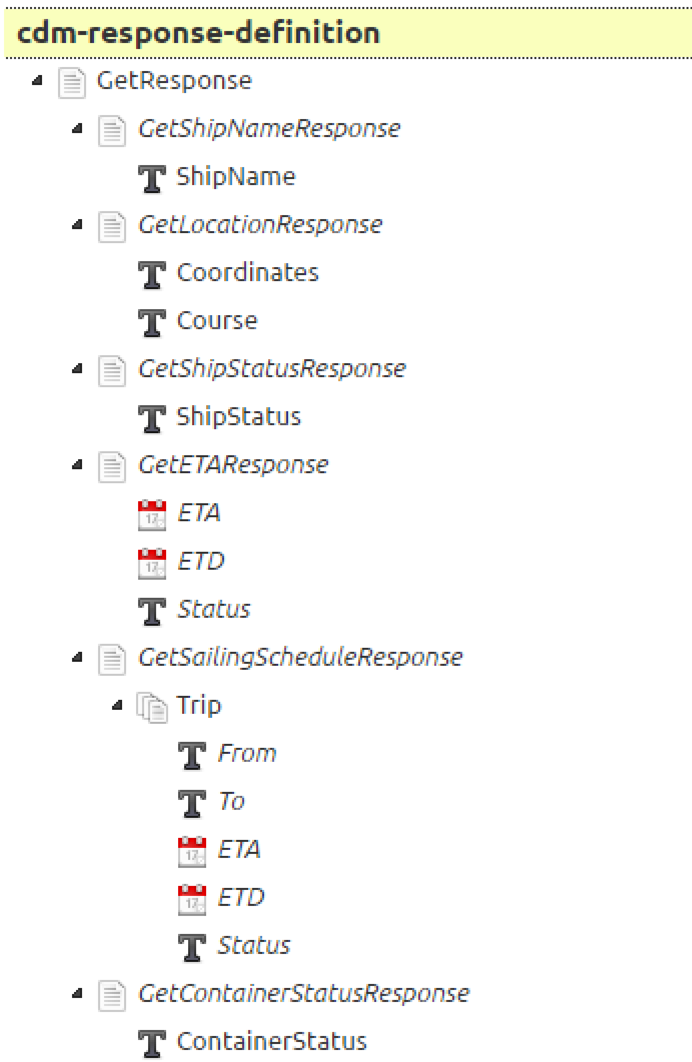

Figure 41: Responses type and the information expected

7.4.5. Issues

Two issues are still open. It concerns the different transformation that is needed within the response flow to transform input messages to a CDM and from a CDM to an accepted output message. When Cloudscrape sends a response it comes in a specific output format. Not only does it send the requested values back to you but it sends back the input values as well. If you take one particular request type, the response values are always the same, but the input values necessary to retrieve this information can be different. Meaning that the response values within the response message can be on different rows. Since the transformations are static, different transformations have to be built when the number of input values differs.

The other issue concerns the Date representation of different websites. Each website has its own notation of a date. This results in the same problem as mentioned above. The transformation is static which means that formatting the time in the desired format for one specific website can be done in one transformation. For multiple date representations, multiple transformations are necessary.

7.5. Cloudscrape

In the architecture, Cloudscrape is placed in the Application Layer shown in Figure 42 and focuses on the Cloudscrape API and the website scraping.

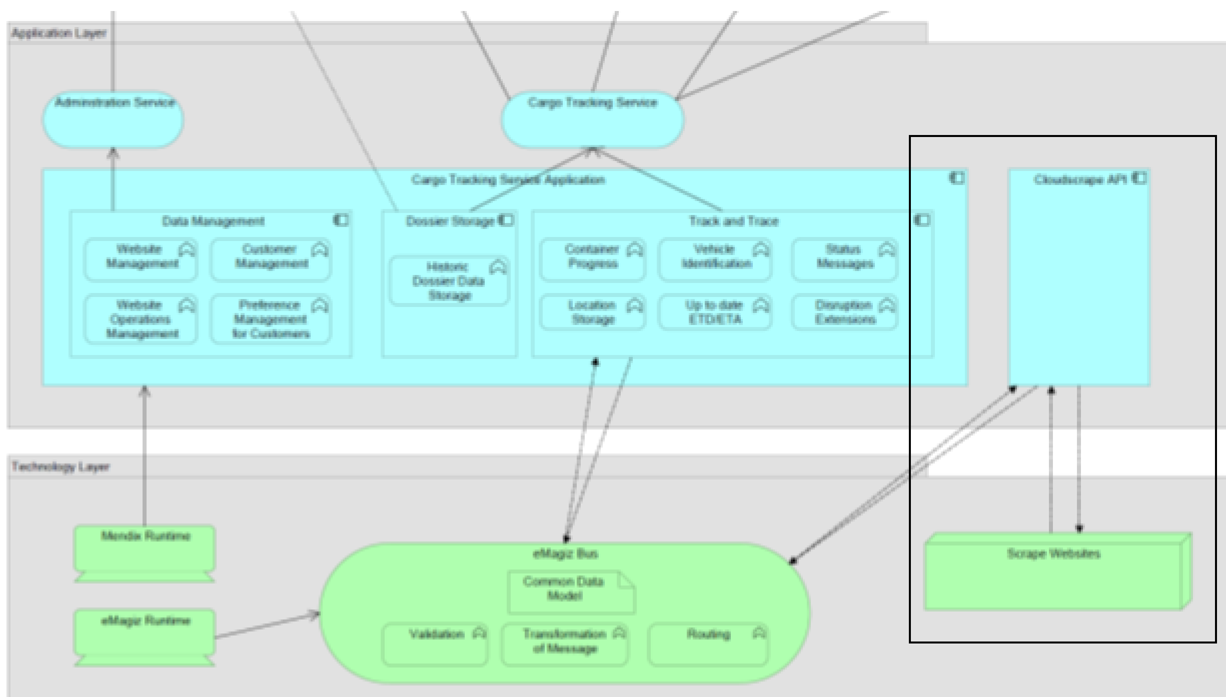

Figure 42: The application and technology overview of the prototype

7.5.1. Cloudscrape Introduction

Cloudscrape is a web scraping tool that scrapes data from websites based on a specific path. It is a browser based application with an easy to understand user interface. The tool offers the possibility to make robots. Robots are designed by the user to follow the actions you go through on a website, to save the data the user want to save and automate this process and send automatically a response when the robot is executed. The selection of data works via CSS selection. Cloudscrape provides a REST API (Representational State Transfer Application Programming Interface) to ease communication with other applications.

7.5.2. Connection to the Research

As described at the research motivation as well at the findings during the interviews and observations, planners visit several websites per shipment in order to retrieve up to date data. The goal is to use the common data model to build a prototype that automatically retrieves this information for the planners and present it to the planner. Automatic retrieval of data is achieved by using Cloudscrape.

7.5.3. Input

In Chapter 3 the information is identified that is needed to constantly update the dossier. The information needed as input to retrieve that information from the website is identified as well. Based on this information the CDM is built and validated in Chapter 5 and 6. During the validation in chapter 6, the information that is needed as input must be retrieved from the order information as they should be retrieved via the back office systems as the Transportation Management System. In the same chapter is validated that the output we needed from the different websites can be retrieved and that the dossier can be updated based on this external data.

7.5.4. Translation to the Prototype

Within Cloudscrape, the user can create robots that scrape websites based on the user's input. An example of how a robot is designed is shown in Figure 43. Based on the findings in Chapter 3, planners retrieve different types of information from one or multiple websites. Each type of information should be processed differently within the Mendix application. Therefore is chosen to split the requests based on the request type. Examples of request types are the retrieval of the current vehicle name, the ETA/ETD or the current location of the vehicle. For every website is identified what types of requests can be requested and these request types are built.

During the design of the robot the semantic conflicts of the data, identified by Zhang, are solved (Zhang, 1994). As shown in figure 43, the robot is designed by steps. Each step indicates an action. One of these actions is the extraction of selected data. During this extraction is identified what type of data the extracted data is, for example, an ETA or ETD.

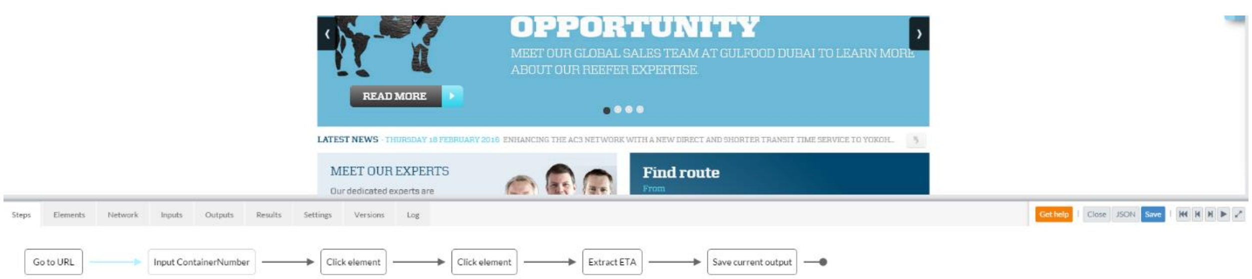

Figure 43: Example Cloudscrape robot

The Cloudscrape API can be called using different methods. Since most robots require inputs and the application want to receive the response as soon as possible, the API method should send the output back in the response as soon as it is done. The method 'executeWithInputSync' is chosen. Using this method the request can send inputs for the robot while the response is sent back immediately when done.

7.5.5. Issues

Often when data are scraped from websites the data you retrieve are not clean. Often the information needed is combined with other words that are not interesting for the information you retrieve. For example, if the ETA is needed, you can scrape only from the website 'The arrival time is: 01/01/2016 21:00'. What the systems needs is only the date and time while the words before the data time are not interesting. With the scraping other HTML elements are scraped as well. Logic is built into the eMagiz bus to transform the sentence in only a date time while Cloudscrape offers the functionality to remove HTML tags. Sometimes this is more difficult and is this a hard issue to solve.

In the Cloudscrape introduction is told that the data are scraped via a CSS Selector. The user is completely dependent on how the owner of the website has organized the websites. Sometimes the layout gives problems when Cloudscrape cannot more complex tables or when a table looks like a table but in the layout is every row a table itself which makes it almost impossible to scrape.

The last issue is that is when data are not specifically labeled on the websites but you need multiple page scrapings to identify the information you want to retrieve. Since it is not specifically labeled on one page, the scraping is a lot harder. It is not possible to put logic in the design of the robot. You can't say, try to find this element and when found, send it back and when not found, go to another page and find the same element.

7.6. Performance of the Prototype

The prototype is the combination of all three applications and services. This section shortly explains how the prototype is tested and what its performance is.

In Chapter 3 is identified that there are a wide variety of websites that can be scraped and that for every website different types of requests can be requested, even the same request type but with different input. To build all these robots for every website and make it work in eMagiz so the Mendix application can receive and process the response is a lot of work, unfortunately too much in the short amount of time of this thesis. Therefore is chosen in consultation with both the university and CAPE Groep to focus on implementing one website for every websites type for one modality. Chosen is for the Deep Sea modality and this has three types of websites: terminal, carrier and AIS data websites. For the terminal website, the APMT 1 Rotterdam Terminal is chosen, for Carrier Maersk and for AIS Data Marinetraffic. For every website, the robots are built and set up and in eMagiz, the transformations are made so the Mendix application can receive the responses.

The application is tested with multiple orders dossiers that fit with the implemented websites. In figure 44 is shown a screenshot of how the prototype looks. In appendix C, multiple screens are shown. Over a period of two weeks, the prototype is tested and bugs are fixed when they occurred before the prototype was stable enough to test. The basis of the prototypes works as expected. Only the two robots in

Cloudscrape could not have been built since the information is not available on the websites and for the other the layout of the websites was an issue. This is also discussed in section 7.5.5.

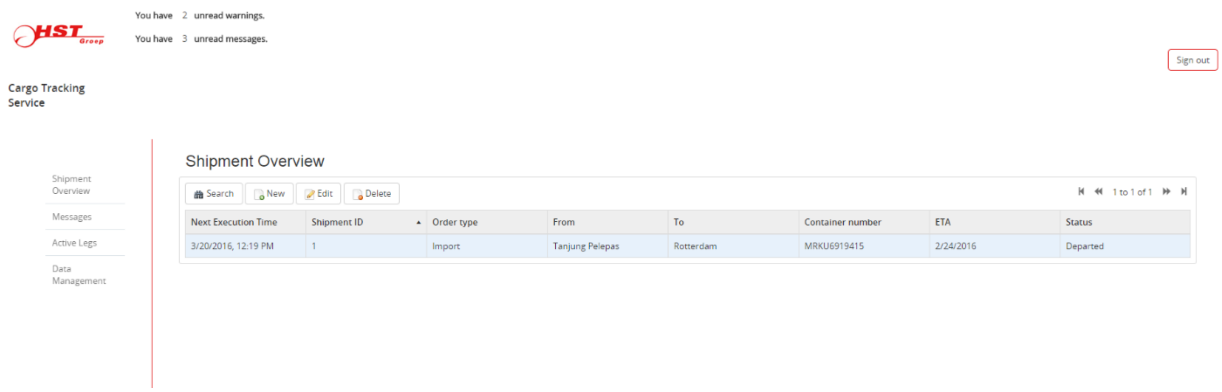

Figure 44: Homepage of the user

The speed of updating one leg is tested on a Monday around 10 AM. Three test runs are executed. The result are shown in table 31.

Table 31: Test run results

Test run	Time
1	8.25 min
2	8.04 min
3	8.31 min
Average	8.20 min

Around 8.20 minutes is necessary for updating one leg but since all legs that need to be updated are triggered simultaneously one can add one or two minutes per leg to update multiple legs. The performance is currently not too good and the prototype is slow resulting from different scripts that are running on the websites. This must be solved when the prototype is tested in a working environment. However, this is tested on open servers of Cloudscrape. Discussions with the CEO of Cloudscrape learned that one can order dedicated servers at Cloudscrape which can lead to a significant increase of performance. This is not tested so no numbers can be given.

Users of the customer use the Cargo Track and Trace service as shown in the architecture while the administrator takes care of adding websites and assign different websites to customers. One application component is not specifically mentioned. The Data Storage components stores all retrieved data in a database which enables the ability to add extra functionalities and information based on data mining.

The prototype proofs that it is indeed possible to scrape data and follow automatic the progress of a shipment. The prototype is not yet bug-free and some error handling is missing. It proofs that for Deep Sea it is possible to track and trace the shipment. Since the model and the prototype is built as generic as possible, one can with less effort extend the prototype with modalities Air and Barge since it is validated in Chapter 6 that all this information is available for both the customer and on the internet.

7.7. Validation of the Prototype

According to the ADR, the prototype has to be validated and evaluated regularly which means that during the design and development the prototype and CDM are evaluated multiple times. Feedback sessions were held with experts from CAPE Groep. These experts have knowledge about the software, the design process, the problem situation and the customers.

This chapter focuses on the evaluation and validation of the CDM and the prototype at the end of the development process when the final evaluation and validation with several logistics providers will be done.

Evaluation means according to Wieringa (Wieringa, 2010) “comparing the objectives of a solution to actual observed results from use of the artifact in the demonstration”. This chapter evaluates two parts of the solution: the common data model and the prototype. Both parts are evaluated on different criteria.

7.7.1. Validation and Evaluation Set-Up

There are many ways to set up validation and evaluation processes using different methods and techniques. This subsection elaborates what validation is used and what the population was.

As validation method, expert opinion (Wieringa, 2010) and a test is held for a period of four weeks at Kuehne-Nagel is chosen. With expert opinion, experts imagine what would happen if the solution would be implemented into the problem context it was designed for.

Five experts were asked, because of their high experience and knowledge about the subject, the problem context and as a potential user. The experts are selected based on the interviews and observations held in Chapter 3. Experts were selected from HST Groep and Kuehne-Nagel. During the project Kuehne Nagel, number 3 of logistics providers in the Netherlands, became very interested in the project and offered their help in this project. The roles of the experts within the companies are shown in table 32.

Table 32: List of Interviewees

Expert Number	Company	Role
1	HST Groep	Planner Import Seafreight
2	HST Groep	Manager Sea-and Airfreight Department
3	Kuehne-Nagel	Innovation Manager
4	Kuehne-Nagel	Site Manager Schiphol Airport
5	CAPE Groep	Process and Data Modelling Expert

7.7.2. Interview Set-up

Interviewing is a method to ask experts about their opinion. Three types of interviews can be defined: structured, semi-structured and unstructured interviews (Bernard, 2006). Semi-structured interviews are most suitable for the expert interviews in this research. They provide the possibility to interview experts about certain topics that are fixed while still being able to go deeper into certain or relating topics (Cohen

& Crabtree, 2006). For this research is chosen to interview all companies separately, not on an individual base. This is chosen so experts from the same company can discuss when necessary while putting all experts together was not achievable since the large distances and fully schedules of the experts.

During the interview, the Mendix application is shown which is the user interface of the platform. Experts have knowledge about the problem situation, what they want and need, but they have no knowledge about data modeling. Since the common data model is regularly validated by an expert within CAPE Groep, the rest of the interviewees have to focus whether all information they need and want is shown in the prototype and whether the prototype is useful, if it will help on a daily basis and it will improve the ability to execute synchromodal modality shifts. The last requirement is derived from the research goal in Chapter 1. All information that is shown in the prototype should be stored in the common data model. If data are missing or irrelevant, this directly involves the CDM.

The interviews were set up in five phases:

1. Introduction
2. Demonstration of the prototype
3. Validation questions about the common data model
4. Validation questions about the prototype
5. Discussion

More detailed information about the interview setup as well as the questions asked can be found in Appendix E.

7.7.3. Prototype Test

During a period of four weeks, the prototype is tested by the inbound department of Kuehne-Nagel at Schiphol. The prototype is used as ETA monitoring system for incoming airplanes. For the first two weeks, the prototype is tested by the head of the planning at the inbound department. During the last two weeks, multiple users tested the prototype. The test is set-up to track the progress of flights and keeping the ETA and the status of the flights up to date.

7.7.4. Evaluation and Validation Results

The answers of the experts during the interviews can be used to validate the designs of both the CDM and the prototype and collect further feedback to improve it. In this subsection, the results of the evaluation and validation are discussed. The elaborated versions of the evaluation and validation interviews can be found in appendix E.

The results show that the interviewees see potential in the designed prototype. The interviewees state that all the information is available to the user, although some suggestions are made to show some information differently. Some suggestions for improvements are made to extend the model and the prototype with customs information as well as including information from the preliminary stage of an order. Another improvement is to extend the personalization functionalities for the user.

Critical feedback suggests that the prototype is useful when it has connections with the back-office systems that will feed the prototype with the information it needs. When a planner needs to put the basic order information manually in the prototype, then the prototype loses its value.

All interviewees estimate that the concept can save money as well as increase the ability to make a more efficient planning. HST Groep estimates that a saving of 700 working hours can be realized yearly. All the interviewed logistics providers are interested in testing this in practice.

The test in practice showed that the prototype saves time in comparison with manually updating. The employees see it as an great advantage that all information is shown in one screen instead of dispersed over several sources. Although it is not tested, all of them see the potential of the prototype to increase data sharing with supply chains partners as well as intern with other departments.

A must-have is that the prototype should update automatically. Unfortunately, it did not because of the Mendix Cloud environment the prototype was running on. This is solved when the prototype runs on another environment. Another con is the speed of updating. Each flight costs around 45 to 60 seconds to update it. An increase in speed is reached when dedicated nodes of the web scraper is used.

7.8 Conclusion

This chapter answers the seventh research question:

SQ7: Has the prototype positive effects on the ability to execute synchromodal modality shifts?

The prototype can be titled as a proof of concept because users should automatically trace the progress of the shipments and is notified when necessary while each of the applications/services has still its unsolved issues, incomplete error handling and it misses a fully connected library of websites to use the prototype on daily basis in practice at the customer. The connection to the customer's back-office system is not made. If this application is will be used on a daily basis at the customer this connection is desirable. It saves a lot of time to automatically retrieve orders when they are accepted in the back-office system. Since those connections do not influence the core functionality of the prototype, which automatically traces the progress of a shipment, this is not included in the prototype.

Evaluation and validation show that the prototype is a successful proof of concept. It is estimated that it can save 700 working hours on a yearly basis while increasing the ability to execute synchromodal shifts since planners are able to make a more efficient planning based on the more accurate data the platform retrieves. A test for four weeks at Kuehne-Nagel has been successful and they see great potential to increase efficiency and data sharing among departments and the supply chain.

Chapter 8 Conclusion

In this chapter, the answers to the research questions will be given. Before answering the main research question the sub-questions will be answered.

This research improves the ability to execute synchromodal modality shifts. Validation interviews with experts in the field show that the prototype with the designed common data model improves this ability. Based on the observations, the research problem is defined as the manual website visits to update shipment that is time-consuming. Designing a common data model and validate this in the working prototype is chosen as solution direction. To define what information must be present in a common data model, the first research questions must be answered.

8.1 How to build an Integration Platform?

The first step in designing a common data model is identifying the problem and requirement definition (Böhmer et al., 2015). Literature is studied to answer four sub-questions 1a-d.

SQ1a: What is the difference of synchromodality in comparison with other types of transportation?

In comparison with other types of transportation, synchromodal transportation gives logistics service providers the possibility to select the carriers or customers independently and, at any time, the best modality based on the operational circumstances and/or customer requirements (Steadieseifi et al., 2014). Riessen et al. state that synchromodality includes the possibility of real-time switching between modalities (Riessen et al., 2013b).

SQ1b: Why do logistics providers need an integration platform?

Literature showed that logistics providers currently often cannot answer simple questions as like ‘‘Where is container now?’’ or ‘‘How long will it take for my shipment to be available in my warehouse?’’ while simple basic information regarding the current status and geographic position of an object in a supply network is often nonexistent or unavailable (Toth & Liebler, 2013). In order to make this information visible to the user, real-time information about the current situation and anticipated future behavior in the supply chain is necessary. Dashboards have the potential to be helpful and successful for logistics (Zaiat et al., 2014).

SQ1c: How is a common data model for logistics designed?

To retrieve information real-time information, Oude Weernink proposes a meta-model of integration platform (Oude Weernink, 2015). Central in the integration platform is the common data model which takes care of the translation of messages from and to different back-end systems and the routing of the messages. Böhmer mentions a scenario for an integration platform and states that a common data model in an integration platform closes the gap between business and IT (Böhmer et al., 2015). It contains four steps: Problem and Requirements Definition, Analysis, Design and Maintenance.

SQ1d: How can big and open data dealt with on the World Wide Web?

Data from the World Wide Web can be structured by information harvesting (Qureshi et al., 2011). With information harvesting structures the data as output, for example in a relational database. Zhang identified different conflicts between the unstructured data on websites (Zhang, 1994). A difference in semantics, schema or data can lead to conflicts as data, semantic and schematic (Zhang, 1994). Web harvesting in combination with transformations in the bus integration system, using a common data model, can solve these conflicts (Poulovassilis & McBrien, 1998). Transformations of messages into the end system or common data model format take care of the harmonization of the data from the World Wide Web. During transformation, the mapping of information translates the conceptual components to the logical components (Domínguez et al., 2005).

8.2 Data Sources

The second step consists of the analysis of the considered domain (Böhmer et al., 2015) and will be answered in the next sub-questions. These sub-questions give insight in what information LSPs use to update their shipments as well what other information is available on the internet what might be interesting to take into account. The question is formulated as follows:

SQ2a. What data sources are used by Logistics Service Providers?

Based on the observations and interviews with planners and manager at Neele-Vat, HST Groep and CTT the necessary data and sources are identified. Currently, the planner uses three different types of data sources to retrieve data for their shipments stored in their back-office systems. For each modality, planners use a selection of these four types based on whether the information is available on these websites for this specific modality. Examples of websites are shown in table 34. The website types are:

- Carrier Websites
- Terminal Websites
- AIS Data Websites
- ADS Data Websites

SQ2b: Which data sources are available?

On the internet, other data sources are available. Several data sources are identified. Carrier and terminal websites are excluded since every carrier or terminal has their own website. Table 33 shows for AIS and ADS the available websites.

Table 33: Identified additional websites on the World Wide Web

Website Type	Website
AIS	www.aishub.net
	www.vesselfinder.com
	shipfinder.co
ADS	www.flightview.com
	planefinder.net
	www.flightaware.com

It turns out that the combination of these four types contains enough information to update a leg with data from websites. However, for the extensibility of the platform, many more websites can be used to retrieve information that might be relevant for forecasting. Information as weather information, traffic information specifically for truck transport or information about the open and closing times of locks at inland waters for barge transport can have effects on the arrival time of the shipment. Such information can be bundled and made automatically available to the planner with the platform which was manually too time-consuming to retrieve. This information is currently nice to have for planners and has no high priority.

Appendix B more detailed information is given about the different modalities, processes and information.

8.3 Data Representations

Data are necessary to say something about a shipment. These data must be retrieved from the different websites. However between different websites, multiple notations and representations of the same data type exist. Zhang identified three different conflicts: schematic, semantic and data (Zhang, 1994).

SQ3. What are the similarities and differences between the different representations of the retrieved data?

During the interviews is asked and identified what information planners need at the different website types in order to take decisions. Based on this information multiple websites per website type are researched in order to identify whether this necessary information can be found and extracted by Cloudscrape (from April 206 dexi.io). In total 14 websites are identified. Table 34 shows which websites are identified. Between modalities is identified that there is an open data inequality. This means that for modalities Deep Sea and Air more than enough data are publicly available while for Barge is much fewer data are available.

Table 34: Identified websites

Website Type	Website	Website Type	Website
Carrier	www.maersk.com (Maersk)		http://ecom.kline.com (K-Line)
	www.afklcargo.com (KLM/Air France)		http://www2.nykline.com/ (Nyk Line)
	www.qrcargo.com (Qatar Airways)	**Terminal**	eservices.ect.nl/ (ECT Terminal)
	www.yml.com.tw (Yang-Ming)		www.apmtrotterdam.nl (APMT 1 Terminal)
	tang.csair.com (China Southern Airlines)	**AIS**	www.marinetraffic.com
	www.turkishcargo.com.tr (Turkish Airlines)	**ADS**	www.flightradar24.com
	cargo.virgin-atlantic.com (Virgin Atlantic Airways)		

The analysis confirmed that between websites semantic and data conflicts exist. Semantic conflicts are caused by the different semantic meanings of the same data. An example of a semantic conflict is different meanings of a date and time. ETA and ETD have different meanings while using date and time. Data conflicts occur when data is represented in different scales, precisions or formats. Again a data and time are taken as an example. Multiple websites show the ETA of a ship and the planner that has to decide which ETA is most reliable. However between these websites, the same data can be presented in a different format. As an example, an ETA value is retrieved from both Marinetraffic as Maersk. At Marinetraffic, a date is presented as: 2016-01-01 00:00. For Maersk, this is: 01 Jan 2016, 00:00. Zhang states these conflicts can be solved by the mapping to the common data model (Zhang, 1994). The common data model makes sure that the information covered is treated unambiguously and can be presented to the user in the same format for each modality.

8.4 Data storage and a more efficient supply chain

As mentioned in section 8.1, the second step consists of the analysis of the considered domain (Böhmer et al., 2015) which includes the identification of what data is necessary for the planners to update their shipments and should be stored in a platform and what data should be shared over the supply chain.

SQ4. What historical and current data should be stored and shared with the rest of the supply chain in order to make it more efficient?

During the interviews and observations is identified what data are necessary to update a shipment. For each modality in the scope and each type of transport (import, export or inland) is identified what data is necessary for the planners. The results, without duplicates, is shown in table 35.

Table 35: Information necessary to update shipments for all modalities

Flight Schedule	ETD	Flight Number	Flight Status
Container Status	Geographical Position	ETA	Ship Name
Ship Status	Sailing Schedule	Course Ship	Container Number
Flight Number	AWB Number	BL/Number	

The results showed in table 35 are just a part of the information that needs to be stored in the prototype for a shipment. Order information about a shipment is necessary from the back-office systems. This information is the basis of the shipment and serves as input for the data retrieval. Zaiat et al. mention that the representation of this data in a dashboard has great potential to be helpful and successful (Zaiat et al., 2014). The integration platform should store all historical data. All this data can be used for data mining processing and be used to define and calculate KPI’s for carriers and terminals.

Data sharing creates a more efficient supply chain. Logistics providers can arrange other partners for further transportation when for example the container arrives at the terminal because of a more accurate ETA and faster status notifications. Therefore, other parties in the supply chain can be integrated into the platform. This enables an important feature of the platform is that partners can easily get a personal page with an overview with for them important information about the coming shipments and their own shipments.

8.5 An Architectural View

Still in the second step of Böhmer, the impacts of the integration platform for the logistics providers will be identified. An enterprise architecture, based on the language Archimate, shows the impact of the integration platform on a business perspective as well as an IT perspective. The architecture is modeled by the method of Iacob et al. (M. E. Iacob et al., 2013).

SQ4: What does an architecture of a supply chain look like from a synchromodal perspective?

The first two steps of the method of Iacob et al. consists of the exploration of the domain of the scope and in the information system landscapes of the collaborating companies. This information is retrieved by the previous sub-questions as well as the interviews. The third step is designing a current and target architecture by modeling the business and information system landscape. By performing a gap analysis the impact of the integration platform can be determined. The architectures and gap analysis are shown in Appendix C.

The gap analysis shows that the on a business process perspective the shipments are not manually updated by the planners anymore but automatic arranged by the track and trace service. Users do not use multiple sources for their information but the integration platform provides the user a single source for the same information. The implementation has several advantages for the logistics provider. Automatically track and tracing shipments not only saves time by removing manual actions, it also increases communication with carriers, customers and other parties in the supply chain. The application enables users to store more detailed information about the shipment than the TMS allows and show this information in different dashboards.

The architecture implements the synchromodal perspective by increasing the optimal, flexible and sustainable usage of different modalities of transportation. The retrieval and creation of visibility of real-time data enable planners to make better decisions but it also saves time since planners use one source for their information, which is shown in one application, and have less contact with their customers about customer's containers. Customers can directly access their information via their own portal in the integration platform.

Based on Bishr's levels of interoperability the implementation of the integration platform in the target architecture the interoperability between systems will be at level 5 where the connected systems are provided with a single global model (common data model) that is an abstraction of all the underlying remote databases (data models of the systems) (Bishr, 1998). Bishr mentioned that the systems can communicate seamlessly, knowledge on their semantics is still required to reach interoperability. Transformations within the integration bus create interoperability.

8.6 The Common Data Model

By answering the first four sub-questions the first and second step of the method of Böhmer are answered (Böhmer et al., 2015). The third step is the design of the common data model. The goal of the common data model is to store planned and actual information, order and status/disruption information into a generic as a possible data model.

SQ6: How is a common data model for planned and actual information, orders and status/disruptions designed?

From the interviews and observations can be derived that there are three types of information for every modality, shown in figure 45:

- Static Order Data: Data that are mainly retrieved via the back-office systems, define the order in the platform and changes seldom. This data are needed to retrieve the data from the websites.
- Semi Real-Time Data: Data that are retrieved from carrier and terminal websites and covers data on a leg level. The data are updated a few times a day.
- Real-Time Data: Data that are retrieved from ADS, AIS and other sources with real-time information (mainly used for additional disruptions). The data are updated regularly and is on a trip specific level.

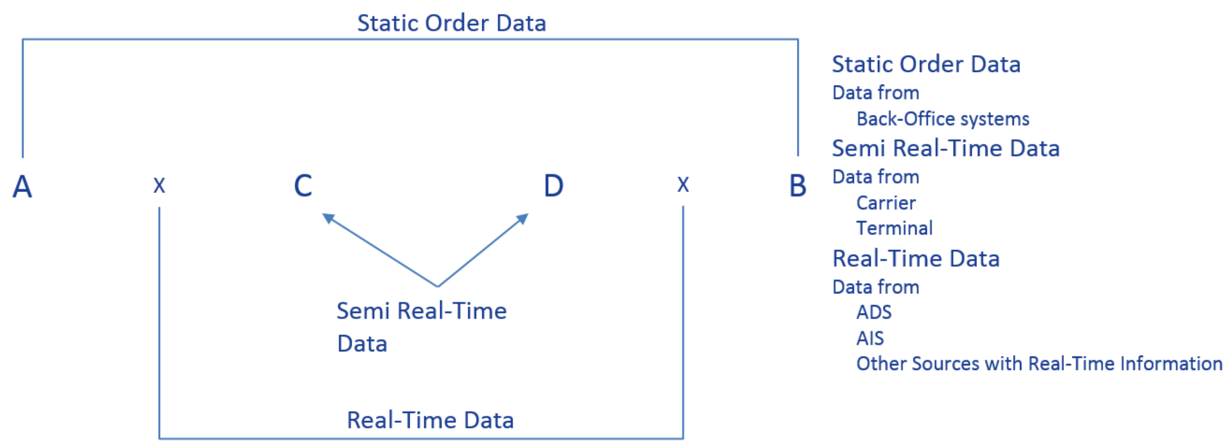

Figure 45: Data types

The design of the CDM consisted of multiple steps. First, a list of data is created with all the information what should be in the CDM and should hold all the information necessary for modalities Deep Sea, Barge and Air. The information necessary from the websites for updating the modalities is shown in table 35. This information combined with the information that is necessary to retrieve the data in table 35. Additional data are added to create the functionality of sending messages and disruption extensions. The second step is the definition of entities that have one or multiple data types as attributes. The last step is to connect the entities and creating relationships between the data. The designed CDM is shown in Figure 46.

The three data types are represented in the CDM. In the red square, all the information related to an order/shipment is stored. The yellow square covers all the data on a leg level while the blue figure covers the real-time data related to a trip level. The CDM is set up as generic as possible which means that each modality is as unambiguously as possible. Only for vehicles differentiation is made. Different modalities use different types of vehicles for transportation. Sea freight uses a ship while air uses an airplane. The different vehicle types have their own unique identifier. For sea freight is the ship name important while for airfreight the flight number is much more important than the unique airplane number. This information will be used to retrieve vehicle specific information as its travel schema or location.

The CDM is validated with four different methods: Bottom-up, Industry Standard, Website Validation Approach and the CAPE projects approach. The Bottom-up Approach validates the CDM by testing the order definitions of the three companies for their applicability with the designed CDM. In the Industry Approach, industrial definitions as IFCSUM and IFTMIN are mapped to the CDM. During the Website Validation Approach, it is tested for every modality if the information that the dossier needs is available on the related websites. At last, an internal project is used to map the information stored in the CDM to the data model of this project.

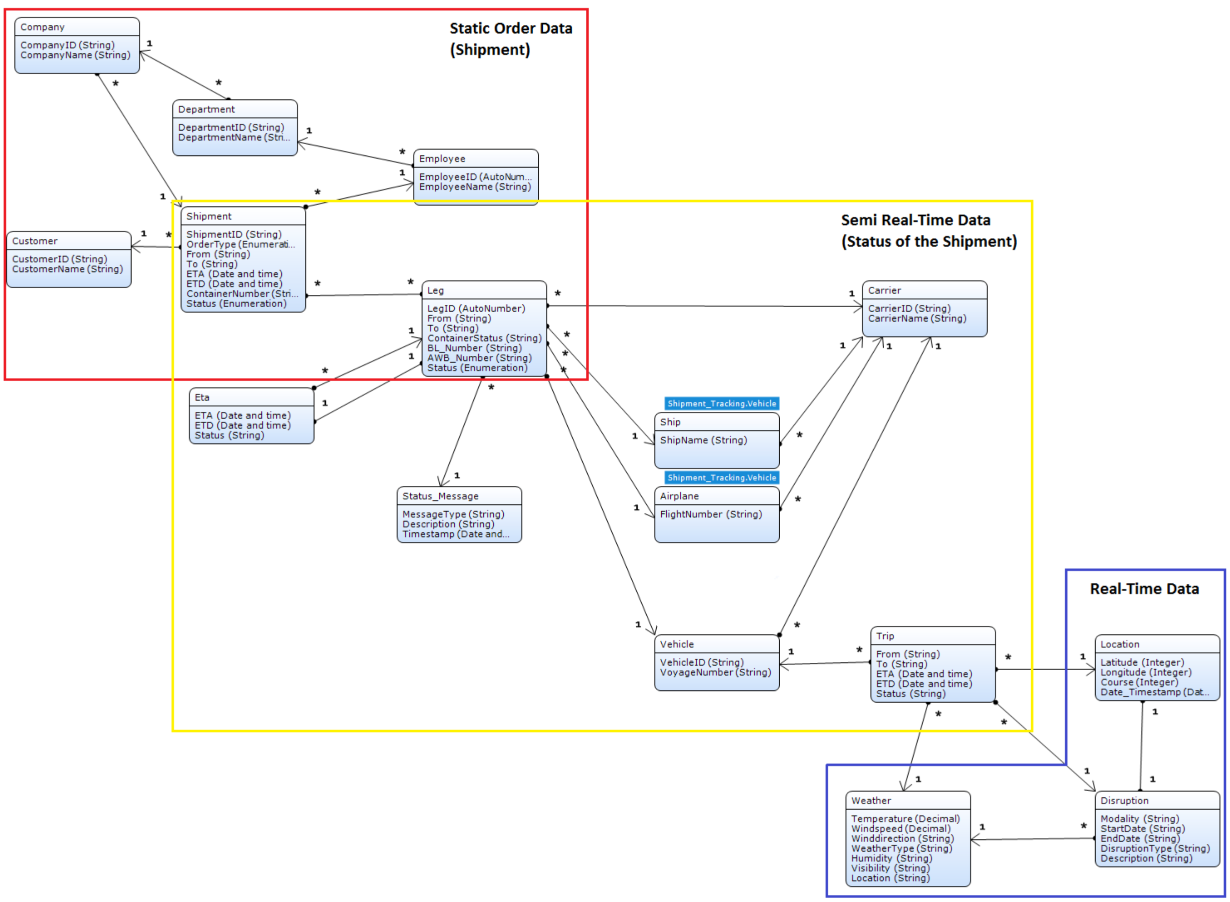

Figure 46: Designed common data model

The CDM tested completed although not all the information can be mapped from various messages, the basic information is available for track and tracing while all the information on the websites is available. The results are shown in table 36.

Table 36: Applicability of the definitions, websites and projects mapped

Approach	Name	Applicability
Bottom-up Approach	HST	+
	Neele-Vat	++
	CTT	--
Industry Standards Approach	IFTMIN	++
	IFCSUM	++
	GS1 NL	+/-
Website Approach	Websites	+
CAPE Projects Approach	SEACON	-

8.7 Development and Testing of the Prototype

Section 8.6 validates the CDM with different definitions, websites and projects to check whether the information is available or can be mapped. The CDM is not validated in a working integration platform. To test the completeness of the CDM a prototype of the integration platform with pluggable service must be designed and tested that implements the models with an application that visualizes the information for the user.

SQ7: How can a prototype be designed implementing the common data model?

In section 8.5 the target architecture showed three layers in the integration platform. The first layer is the application layer, which is an application that creates visibility to the user and processes the data. The application, designed and created in Mendix, sends requests to the integration bus, the second layer. The integration bus is the integration layer and is responsible for the routing and transformation of messages. For the integration bus eMagiz is used as a platform. The request is routed by the bus to Cloudscrape, the web scraping tool. This is the third layer. Cloudscrape scrapes the requested data from the internet and sends the response back to the bus. The conceptual design is shown in figure 47. It includes the Service and Back-office connections as well. These are as well mentioned in the architecture but are excluded from the scope of the prototype.

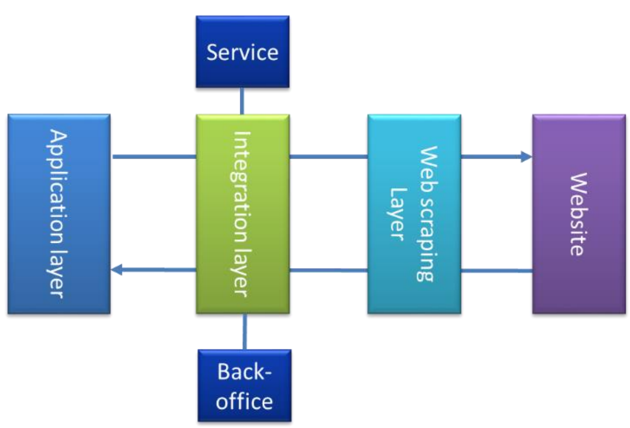

Figure 47: Conceptual design of the prototype

The Mendix application is the user interface of the prototype and has three main functionalities:

1. Track and trace functionalities, includes updating the shipments by requesting data and process this by business logic.
2. Dossier storage functionalities, includes the storage of all data in the application.
3. Data management functionalities, includes the management of personalization features for logistics providers and customers.

In the eMagiz platform, the integration bus is built. Its primary functionalities include the routing, transformation of the messages and as extra functionality the validation of the message. It validates the message based on content and structure. The bus is built as generic as possible resulting that every request has the same format. This created the flexibility to use this request not only in the current context of the platform but makes it useful for all requests to Cloudscrape.

The Cloudscrape service is the web scraping tool used. It scrapes data from websites based on a specific path defined by the designer of the web scraping robot. For most scraping requests input is required. This is retrieved via the Mendix application that sends a request with the necessary inputs. eMagiz requests data via an API and gets an immediate response when the robot is finished. The prototype is tested with one sea freight dossier over several weeks using information from carrier, terminal and AIS data websites. During the testing period, errors and bugs are found and solved. The validation and evaluation of the prototype answer the last sub-question.

> *SQ7. Does the prototype have positive effects on the ability to execute synchromodal modality shifts?*

Validation and evaluation with HST Groep and Kuehne+Nagel is done via semi-structured interviews and resulted in a list of feedback and suggestions for improvements. All the interviewees were excited in the platform and saw multiple uses in their own business cases. Expectations of HST Groep are that on a yearly basis 700 working hours can be saved while the information presented and the messages and warnings send to the planners create a positive effect on the ability to execute synchromodal modality shifts. Instead of having one updating moment a day, they can with the platform create multiple updating moments each day and be updated faster about the process of the shipment. The faster the update is received, the better the planner can make decisions and plan more efficient modality shifts.

Although no measurements are done in a working environment, the expectations are, based on the feedback, that the prototype has the potential to have positive effects on the ability to execute synchromodal modality shifts. HST Groep estimates a yearly saving of 700 working hours while Kuehne-Nagel has successfully worked with the prototype over a period of four weeks.

8.7 Conclusion of this Research

Using the answers to all subquestions, it is possible to answer the main research question:

> *RQ: What is a common data model for logistics for planned and actual information, orders, statuses and disruptions?*

The literature study shows that a need for an integration platform exists (Oude Weernink, 2015; Toth & Liebler, 2013; Vivaldini et al., 2012; Zaiat et al., 2014). Much information is currently manually retrieved from the internet but no platform exists yet that can cope with the difficulties open data have and present the data in an unambiguous format. Difficulties as semantic data interoperability must the platform deal with since the data must be presented to the planner in one format (Zhang, 1994). Researchers indicate the importance of the common data model in an integration platform (Böhmer et al., 2015; Oude Weernink, 2015; Zaiat et al., 2014). A common data model is central to an integration platform and expresses all objects that are available in the different schema's of the connected systems and solves interoperability conflicts (Singh & van Sinderen, 2015; Zhang, 1994).

Interviews and observations were held to identify processes and the information that is needed for updating shipments and must be included in the common data model. Three types of data are identified, differentiated by the frequency that the data are changed. These types are: Static order data, Semi real-time data and Real-time data.

To understand the impacts at the logistics providers architectures are designed showing that by the integration of the platform the business processes will be simplified and that the time the user uses per dossier decline since they only have one source of information instead of many websites and direct communication with the customer over phone or email declines. The integration requires that the planners need access to the platform in order to retrieve the information needed. The architecture shows three layers between the user and the data on the websites. The layers combined are the basis of the prototype. The Mendix application processes the data and presents it to the user. It runs in the cloud or on the customer's infrastructure. Mendix communicates with Cloudscrape via an integration bus to retrieve the data necessary. The integration bus is built in eMagiz and is responsible for the routing and transformation of the messages. Cloudscrape is the web scraping tool that retrieves the requested data from the websites.

The common data model is designed based on the method of Böhmer (Böhmer et al., 2015) and combines the data that planners need from websites with the data that is necessary from back-office systems as input to retrieve the data from websites and data that is necessary to organize the shipments. The results of the common data model are shown in figure 48. The common data model is designed independently of the collaborating logistics provider meaning that it can be applied by other logistics providers.

The CDM is validated using approaches with LSPs order definitions, industry order standards, websites and data models from internal projects of CAPE Groep. The CDM is not applicable to all the data models and standards used in the validation of the CDM. Some message standards or the data model from the internal project miss crucial information that is necessary to retrieve all the necessary data from the internet. The applicability is of the researched definition, standards and projects are shown in table 36.

The prototype is a proof of concept showing that the CDM is useable for its intended purposes. Validation of the prototype shows that experts expect a decline in costs while being able to plan more efficient and thus be able to increase the ability to execute modality shifts. Based on this feedback it can be concluded that the implementation of the integration platform, with the designed common data model, narrows the gap between business and IT and solves the research problem.

The contribution of this thesis is the designed common data model and will help the supply chain with sharing data, the logistics providers to increase data sharing with their customers and planners to save time and presenting unambiguously up to date data for better decision making. The common data model covers planned and actual information, orders, statuses and disruptions and harmonizes unstructured and structured data by presenting the data unambiguously to the user for every modality in the scope of this thesis. The common data model used in the integration bus fosters communication between different parties in the supply chain. Supply chain partners can be integrated into the platform while before direct communication occurred between two partners. In the platform, every connected partner can communicate with each other. This will lead to an increase in data shared among the supply chain while the users can use more real-time information as well. This combination will lead to an increase in the ability to execute synchromodal modality shifts among the whole supply chain.

Suggestions on how to improve the CDM and the prototype will be given in the next chapter, as part of the ‘recommendations for practice’.

Chapter 9 Recommendations

Based on the results several recommendations for science can be provided, as well recommendations for practice. For science suggestions for future research are provided in section 9.1, whereas suggestions for practice are given in section 9.2.

9.1 Recommendations for Science

The literature study shows that multiple researchers recognize the need for logistics providers for an integration platform that is fed with open data that supports users in their decision process (Oude Weernink, 2015; Toth & Liebler, 2013; Zaiat et al., 2014). This section suggests several recommendations for research topics for science.

Firstly, this research has not tested the integration platform in a work setting with a high amount of orders to measure performance and the actual impact on the business processes of the logistic processes. More in-depth research is required in the economics of synchromodality as a result of the improvements of this integration platform.

Secondly, web scrapers scrape data from a website according to a defined path by the designer. Currently, web scrapers cannot cope with changes in the layout of the website. For each change in the layout, the designer must manually adjust the path. An interesting research topic is a focus on how web scrapers can learn to cope with layout changes of websites and automatically change the path.

Thirdly, more in-depth research into the applicability and the impact of the common data model and the integration platform can be done to generalize the results of this thesis. Therefore, the integration platform and common data model must be tested and measured at a number of logistics providers active in a number of different sectors.

Finally, the scope of the common data model can be much wider. Other stages during the order process can be added. Research can be done to identify whether it is interesting to add the planning stage to the CDM. Carriers have not only actual trip information but have planning information available as well. The scope can also be extended with the customs information. All import goods have to go through customs before it can transport further. Other stages might be interesting as well.

9.2 Recommendations for Practice

The prototype showed that the implemented CDM satisfies the requirements of its intended use. Nevertheless, this prototype should be tested in an operational setting in a larger test, letting planners use the prototype and quantitatively assess its performance. More scenarios will be tested when a high amount of shipments are tracked. With more testing, the synchromodal ability can be measured.

Many data are publicly available on the internet, although some can concern about the quality and reliability of the data. Often when a website updates the information on their website, no timestamp is given. This makes it hard to interpret whether the data are up to date. Also, more research must be done to deal with fuzzy and noisy data.

In this thesis, one disruption was added to the CDM to show the extensibility of the data model. Research can be done to specify what other disruptions are important in practice for each modality and whether this data is publicly available. Examples of disruptions are the tides and lock times for barge or weather information.

The platform collects information and stores it in the platform. An interesting research topic for CAPE Groep is to research what can be done with all this historical data. All this data can hold important information that can be derived via data mining. For example, can the performance of carriers and vehicles be calculated based on KPI’s or can the routes of vehicles be calculated based on the location data. Especially combining this with the previous improvement, may lead to predictions of disturbances.

References

Abiteboul, S., Buneman, P., & Suciu, D. (1999). *Data on the Web: From Relations to Semistructured Data and XML (The Morgan Kaufmann Series in Data Management Systems)*. *Database Theory Column*. Retrieved from http://www.amazon.co.uk/Data-Web-Relations-Semistructured-Management/dp/155860622X

ALICE WG2. (2014). Corridors , Hubs and Synchromodality Research & Innovation Roadmap, (December). Retrieved from http://www.etp-logistics.eu/cms_file.php?fromDB=11630&forceDownload

Baranowski, L. (2013). *Development of a Decision Support Tool for Barge Loading*.

Baumgartner, R., Gatterbauer, W., & Gottlob, G. (2009). Web data extraction system. In *Encyclopedia of Database Systems* (pp. 3465–3471). Springer.

Bernard, H. R. (2006). *Interviewing: Unstructured and Semistructured*. *Research Methods in Anthropology*. doi:10.1525/aa.2000.102.1.183

Bishr, Y. (1998). Overcoming the semantic and other barriers to GIS interoperability. *International Journal of Geographical Information Science*, *12*(October 2014), 299–314. doi:10.1080/136588198241806

Blumberg, R., & Atre, S. (2003). The Problem with Unstructured Data. *DM Review*, *13*, 42. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=10186785&lang=pt-br&site=ehost-live

Böhmer, M., Schmidt, M., & Weißenberg, N. (2015). Cloud Computing for Logistics. In M. Hompel, J. Rehof, & O. Wolf (Eds.), (pp. 77–117). Cham: Springer International Publishing. doi:10.1007/978-3-319-13404-8_6

Booch, G. (1998). *The Unified Modeling Language*. *Techniques*.

Brandt, P., Basten, T., Stuijk, S., Bui, V., Clercq, P. De, Pires, L. F., & Sinderen, M. Van. (2013). Semantic interoperability in sensor applications Making sense of sensor data.

Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., & Yergeau, F. (2008). Extensible Markup Language (XML) 1.0, Fifth Edition. Retrieved October 13, 2015, from http://www.w3.org/TR/xml/

Britton, C., & Peter, B. (2004). *IT Architectures and Middleware*.

Cabanillas, C., Baumgrass, A., Mendling, J., Rogetzer, P., & Bellovoda, B. (2014). Towards the Enhancement of Business Process Monitoring for Complex Logistics Chains, *171*, 305–317. doi:10.1007/978-3-319-06257-0

Chen, P. P.-S. (1976). The entity-relationship model---toward a unified view of data. *ACM Transactions on Database Systems*, *1*(1), 9–36. doi:10.1145/320434.320440

Chung, S. M., & Mah, P. S. (1995). Schema integration for multidatabases using the unified relational and object-oriented model. *Proceedings of the 1995 ACM 23rd Annual Conference on Computer Science - CSC '95*, 208–215. doi:10.1145/259526.259556

Cohen, D., & Crabtree, B. (2006). Semi-structured Interviews Recording Semi-Structured interviews. *Qualitative Research Guidelines Project*, 2. Retrieved from http://www.qualres.org/HomeSemi-

3629.html

Domínguez, E., Lloret, J., Rubio, Á. L., & Zapata, M. A. (2005). Evolving XML schemas and documents using UML class diagrams. In *Lecture Notes in Computer Science* (Vol. 3588, pp. 343–352). Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-26844554069&partnerID=40&md5=0a05ebc5f2e855ce20315d16dc86d8de

Drouilhet, P. R., Knittel, G. H., & Orlando, V. A. (1996). Automatic Dependent Surveillance Air Navigation System.

Egberink, J. (2015). *THE INFLUENCE OF TRUST ON INTER- ORGANIZATIONAL INFORMATION SHARING IN LOGISTIC OUTSOURCING RELATIONSHIPS*.

European Container Terminals. (2011). *De toekomst van het goederenvervoer*.

Forum, I. T. (2010). *Illustrated Glossary for Transport Statistics*.

Golinska, P., & Hajdul, M. (2012). Virtual Logistics Clusters – IT Support for Integration Introduction - Virtual Logistics Clusters. *Intelligent Information and Database Systems*, 449–458.

Groep, H. (2015). History - HST Groep. Retrieved September 30, 2015, from http://www.hst.nl/en/about-hst-group/history/

GS1. (2007). Logistics Interoperability Model (LIM), (1). Retrieved from http://www.gs1.org/sites/default/files/docs/sectors/transportlogistics/LIM_Foundation_Report.pdf

Hendrik Haan, G., van Hillegersberg, J., de Jong, E., & Sikkel, K. (2013). Adoption of wireless sensors in supply chains: A process view analysis of a pharmaceutical cold chain. *Journal of Theoretical and Applied Electronic Commerce Research*, *8*(2), 138–154. doi:10.4067/S0718-18762013000200011

Iacob, M. (2013). Sct architectuur, (April), 1–60.

Iacob, M. E., Van Sinderen, M. J., Steenwijk, M., & Verkroost, P. (2013). Towards a reference architecture for fuel-based carbon management systems in the logistics industry. *Information Systems Frontiers*, *15*(5), 725–745. doi:10.1007/s10796-013-9416-y

Iacob, M.-E., Jonkers, H., Quartel, D., Franken, H., & van den Berg, H. (2012). *Delivering Enterprise Architecture with TOGAF and Archimate*.

Kanaoka, K., Fujii, Y., & Toyama, M. (2014). Ducky : A Data Extraction System for Various Structured Web Documents, 342–347.

Krant, L. (2015). Top 100 logistiek Diensverleners 2015. Retrieved September 30, 2015, from http://www.nvc.nl/userfiles/files/Logistiek_Tabel_Top100_2015_DEF.pdf

Lampathaki, F., Mouzakitis, S., Gionis, G., Charalabidis, Y., & Askounis, D. (2009). Business to business interoperability: A current review of XML data integration standards. *Computer Standards & Interfaces*, *31*(6), 1045–1055. doi:10.1016/j.csi.2008.12.006

Li, L., Negenborn, R. R., & De Schutter, B. (2013). A general framework for modeling intermodal transport networks. *2013 10th IEEE INTERNATIONAL CONFERENCE ON NETWORKING, SENSING AND CONTROL (ICNSC)*, 579–585. doi:10.1109/ICNSC.2013.6548803

Liam, Q. (2015a). Transformation - W3C. Retrieved October 13, 2015, from http://www.w3.org/standards/xml/transformation.html

Liam, Q. (2015b). XML Schema - W3C. Retrieved October 13, 2015, from http://www.w3.org/standards/xml/schema

Logistics, N.-V. (2015). Neele-Vat Logistics. Retrieved September 3, 2015, from http://www.neelevat.nl/over-ons/

Lucassen, I., & Dogger, T. (2012). Synchromodality Pilot Study - Identification of bottlenecks and possibilities for a network between Rotterdam, Moerdijk and Tilburg.

Mangisengi, O., Huber, J., Hawel, C., & Essmayr, W. (2001). A Framework for Supporting Interoperability of Data Warehouse Islands Using XML, 328–338.

Nations, U. (1980). United Nations Conference on a Convention on International Multimodal Transport United Nations Conference on a Convention on International Multimodal Transport, *I*(May).

Niemi, E. (2006). Enterprise Architecture Benefits : Perceptions from Literature and Practice, 14–16.

NMFTA. (2015). National Motor Freight Traffic Association. Retrieved September 24, 2015, from http://www.nmfta.org/pages/scac?AspxAutoDetectCookieSupport=1

Oude Weernink, M. J. P. (2015). Development of an Integration Platform Metamodel, 1–10.

Ouksel, A., & Sheth, A. (1999). Semantic Interoperability in Global Information Systems. A brief Introduction to the Research Area and the Special Section. *SIGMOD Record*, *28*(1), 5–12. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Semantic+Interoperability+in+Global+Information+Systems+A+brief+introduction+to+the+research+area+and+the+special+section#1

Pedersen, J. T., Paganelli, P., Knoors, F., Meyer-Larsen, N., & Davidsson, P. (2011). One common framework for information and communication systems in transport and logistics. Retrieved from http://novalog.eu/wp-content/uploads/2012/10/Common-Framework-report-final.pdf

Pleszko, J. (2012). MULTI-VARIANT CONFIGURATIONS OF SUPPLY CHAINS IN THE CONTEXT OF SYNCHROMODAL TRANSPORT, *8*(4), 287–295.

Poulovassilis, A., & McBrien, P. (1998). A general formal framework for schema transformation. *Data & Knowledge Engineering*, *28*(1), 47–71. doi:10.1016/S0169-023X(98)00013-5

Qureshi, P. A. R., Memon, N., Wiil, U. K., Karampelas, P., & Sancheze, J. I. N. (2011). Harvesting Information from Heterogeneous Sources. *2011 European Intelligence and Security Informatics Conference*, 123–128. doi:10.1109/EISIC.2011.76

Riessen, B. Van, Negenborn, R. R., Dekker, R., & Lodewijks, G. (2013a). Analysis of impact and relevance of transit disturbances of planning in intermodal container networks, 1–22.

Riessen, B. Van, Negenborn, R. R., Dekker, R., & Lodewijks, G. (2013b). Service network design for an intermodal container network with flexible due dates/times and the possibility of using subcontracted transport. *Report / Econometric Institute, Erasmus University Rotterdam*, (June 2013), 1–16. doi:10.1504/IJSTL.2015.069683

Runhaar, H. (2001). Efficient Pricing in Transport. *European Journal of Transport and Infrastructure*

Research, *1*(1), 29–44.

Sein, M. K., Henfridsson, O., Rossi, M., & Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, *35*(1), 37–56.

Sheth, A. P., & Larson, J. a. (1990). Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys*, *22*(3), 183–236. doi:10.1145/96602.96604

Singh, P. M., & van Sinderen, M. (2015). Interoperability challenges for context-aware logistic services - The case of synchromodal logistics.

Sliwczynski, B., Hajdul, M., & Golinska, P. (2012). Standards for Transport Data Exchange in the Supply Chain – Pilot Studies. *Agent and Multi-Agent Systems. Technologies and Applications*, 586–594.

Steadieseifi, M., Dellaert, N. P., Nuijten, W., Van Woensel, T., & Raoufi, R. (2014). Multimodal freight transportation planning: A literature review. *European Journal of Operational Research*, *233*(1), 1–15. doi:10.1016/j.ejor.2013.06.055

Tamm, T., Seddon, P. B., Shanks, G., & Reynolds, P. (2011). How does enterprise architecture add value to organisations? *Communications of the Association for Information Systems*, *28*(1), 141–168.

Tan, Y.-H., Klievink, B., Overbeek, S., & Heijmann, F. (2011). The Data Pipeline.

Teorey, T., Lightstone, S., & Nadeau, T. (2006). *Database Modeling and Design*.

TOGAF. (2009). The open group architecture framework, version 9, enterprise edition, Document number: G091. Retrieved October 22, 2015, from http://pubs.opengroup.org/architecture/togaf9-doc/arch/

Toth, M., & Liebler, K. M. (2013). Real-Time Logistics and Virtual Experiment Fields for Adaptive Supply Networks, 327–340. doi:10.1007/978-3-642-30749-2

University of Twente. (2016). Synchromodal IT: IT Services for Synchromodality (DINALOG/Topsectors). Retrieved March 14, 2016, from https://www.utwente.nl/ctit/research/research_projects/national/industry/synchromodal-it.html

van der Burgh, M. (2012). Synchromodal transport for the horticulture industry, (December), 89.

Vasani, K. (2014). Content Evocation Using Web Scraping and Semantic Illustration. *IOSR Journal of Computer Engineering (IOSR-JCE)*, *16*(3), 54–60.

Veldhuis, R. (2015). DEVELOPING AN AUTOMATED SOLUTION FOR ETA DEFINITION CONCERNING LONG DISTANCE SHIPPING, 1–51.

Vivaldini, M., Pires, S. R. I., & Souza, F. B. De. (2012). Improving Logistics Services Through the Technology Used in Fleet Management. *Journal of Information Systems and Technology Management*, *9*(3), 541–562. doi:10.4301/S1807-17752012000300006

Weske, M. (2007). *Business Process Management*.

West, M. (1999). Developing High Quality Data Models. *The European Process Industries STEP Technical Liaison Executive (EPISTLE)*, 62.

Wieringa, R. (2010). *Design science methodology*. *the 32nd ACM/IEEE International Conference*.

doi:10.1145/1810295.1810446

Wijaya, W. M., & Nakamura, Y. (2013). Predicting ship behavior navigating through heavily trafficked fairways by analyzing AIS data on apache HBase. *Proceedings - 2013 1st International Symposium on Computing and Networking, CANDAR 2013*, 220–226. doi:10.1109/CANDAR.2013.39

Zaiat, A., Rossetti, R. J. F., & Coelho, R. J. S. (2014). Towards an Integrated Multimodal Transportation Dashboard.

Zhang, J. (1994). A Formal Speci cation Model and Its Application in Multidatabase Systems 1 Introduction.

Appendix A: Interview Set-up and Interview Summary for Process Identification

The questions asked during the semi-structured interviews are listed. The interviews were conducted in Dutch. The interviews are elaborated on the next pages but are in Dutch. Each interview has a small summary in English that discusses the important results of the interview

At the start of the interview, the setup of the research and the research problem are introduced. After this, experts were asked about:

1. How does the planning process of a shipment look like within your company?
2. Why is it important to track cargo during its transportation?
3. What is the process to update cargo on a regularly basis?
   With an observation of the process.
   a. What systems and websites are used?
   b. How much time does it cost to update one shipment?
   c. What are crucial decision points in the process?
4. What data are necessary to track the cargo and to make decisions when problems occur?
5. What do you want to improve in this current updating process?

After these questions interesting findings are discussed in more detail.

Interview Planner – Import Department – Neele Vat

Held on September 29, 2015

Summary

The interviewee is part of the import section for deep sea shipments. During the interview several important decisions points in the track and trace process are discussed. The vehicle name where the container is location should always be confirmed before retrieving the rest of the shipment information that is mainly based on the vehicle. Within a week of arrival the frequency of updating increases while the data source used shifts from carrier websites to terminal websites. Ideally this information is retrieved automatically while being updated when the data as ETA shifts too much.

Interview

De planner maakt onderdeel uit van de import afdeling van de zeevracht. Zij houden zich vooral bezig met het verzorgen van binnenkomende orders tot de afhandeling van transport in Rotterdam. Het proces begint met het ontvangen van een order om een lading te verschepen naar Rotterdam. Wanneer dit een LCL lading betreft, gaat de agent de orders groeperen totdat een volle container is bereikt. Wanneer dit een FCL is, dan is de import afdeling bepalend voor welk schip gekozen wordt.

Nadat uiteindelijk een order definitief is bevonden en gemaakt, het schip/lading wordt bijgehouden. Allereerst is dit via een schipnaam die bekend is bij de Import, maar wanneer een B/L is ontvangen de containernummer wordt dan leidend. De containernummer wordt gebruikt omdat het nog wel eens voor wilt komen dat een container wordt verwisseld van boot zonder dat de B/L wordt aangepast. Echter de rederij biedt deze informatie wel aan via het containernummer, net als informatie over de status van de container (wel of niet geladen). Tot een week voor aankomst, de containernummer is leidend om een schipnaam te vinden. Via dit schipnaam wordt een vaarschema opgevraagd waar de ETA in Rotterdam is weergegeven. Deze wordt wekelijks aangepast.

Wanneer een schip binnen 7 dagen moet aankomen op Rotterdam, de frequentie van ETA opzoeken gaat naar dagelijks en de gebruikte databronnen verschuift van de rederij naar de terminal websites. Op de terminal websites staat namelijk ook de status van de boot en de ETA. Wanneer de boot is aangekomen, er wordt bijgehouden via het containernummer of de container wel of niet aan land is. Wanneer dit namelijk zo is, de douane kan worden geregeld. Voordat het niet op land is, mag dit nog niet. Uiteindelijk kan de status van de container ook gedeeld worden met de klant.

Het ideaal beeld dat wordt geschetst is dat het dossier automatisch wordt geupdate op basis van het containernummer en dat er wordt aangegeven wanneer er grote verschuivingen in ETA zijn of wanneer de container aan land is.

Een meerwaarde zal ook zijn dat het dossier al aangemaakt wordt wanneer een order definitief is in plaats van dat een dossier wordt aangemaakt als het b/l nummer bekend is. Op deze manier kan ook het voortraject van het schip in de gaten worden gehouden. Hierbij verandert wel het huidige proces van de afdeling.

De import werkt voornamelijk met 8 rederijeren, waarmee ze contracten hebben lopen. De 5 belangrijkste zijn ook nog eens onderstreept:

- K Line
- China Shipping
- Evergreen
- CMA
- MSC
- OOCL
- Maersk
- Nyk Line

Interview Manager – Export Afdeling – Neele Vat Zeevracht

Held on September 29, 2015

Summary

The interviewee is the manager of the export department for deep see. Although at first some miscommunication about the scope of the assignment, several points of interested are discussed. Automatic retrieval of vehicle planning from carriers are interesting to retrieve. At the export, the planner collects information as opening and closing times for the cargo as well as the confirmation that the vehicle has departed. Important is to confirm whether the container is on board or not. However this can be confirmed with the retrieval of the Master B/L, the combination of the departure of the vehicle and that the container is not in the port anymore gives a good indication.

Interview

De geinterviewde is een manager van de export afdeling van de Zeevracht die alle export orders plant en regelt tot daadwerkelijk de container op de juiste schip staat. Het gesprek begint nogal moeizaam omdat hij direct al wilt weten wat voor meerwaarde een eventueel scraping applicatie kan bieden voor hem en zijn afdeling. Nadat hij toch na 10 minuten een klein beetje overtuigd wordt kan het gesprek eindelijk beginnen.

Al snel blijkt dat de scope van mijn opdracht/applicatie niet goed is eeft begrepen. Alles kan volgens hem geautomatiseerd worden: email en document genereren en versturen tot aan het automatisch bepalen van tarieven. Hoewel dit voordelen kan bieden voor de afdeling is dit niet in lijn met de scope van de opdracht.

We krijgen een snelle cursus hoe de export afdeling werkt, maar dit is voor mij niet echt van belang. Uiteindelijk komen er alsnog een viertal punten waarbij de scraping applicatie kan werken om handmatige zoektochten op het internet te automatiseren:

- Het automatisch zoeken van afvaarten bij een bepaalde rederij naar Rotterdam. Bij een binnenkomende order kijkt de afdeling welke rederij/schip de container kan vervoeren binnen de gestelde criteria. Hierbij wordt gezocht naar een afvaart schema naar Rotterdam.
- Het dossier updaten via het automatisch updaten van de closing-time via de terminal websites. De closing time houdt in dat tot die tijd containers aangeleverd kunnen worden. Is een schip vertraagd, dan betekent ook dat de closing time van dat schip in Rotterdam veranderd.
- Belangrijk om te weten is of het schip ook daadwerkelijk vertrokken is.
- Net zo belangrijk is om te weten of de container ook daadwerkelijk aan boord is. Hoewel dit pas 100% zeker is al de Master B/L is ontvangen, kan worden nagegaan of het schip is vertrokken en dat de container niet meer in Rotterdam is. Dit geeft een goede indicatie.

Interview Administrative Employee – Neele Vat

Held on September 29, 2015

Summary

The interviewee works at the customs department and processes the application of the containers. Most important information is the ETA and ETD of the container but for each new application the application must be created in the system. The information needed for this application is retrieved from external sources as mail conversations and carrier and terminal websites. Conclusion of the interview is that customs is currently not in the scope of the assignment but can be added as extra functionalitiy.

Interview

De geinterviewde werkt op de afdeling Douane waar containers door bedrijven worden aangegeven om gecontroleerd te worden door de douane. Hij verzorgt de aanvraag bij de douane en doet het 'papierwerk'.

Bij het binnenkomen van een nieuwe aangifte moet ook in het systeem een nieuwe aangifte worden aangemaakt. Hier wordt handmatig de informatie ingevuld die in de benodigd zijn. Deze worden uit de mail gehaald, dan wel uit terminal websites als ect, apmt, rwg of in het allerslechtste geval de rederij website. Voornamelijk wordt hier data uitgehaald die moet worden gecontroleerd met data uit andere bestanden, waaronder de mail. De status van het schip is ook belangrijk voor het doorsturen van de aangifte naar de douane. Dit mag namelijk alleen wanneer het schip daadwerkelijk is aangekomen op Rotterdam. Het belangrijkste informatie dat wordt gecontroleerd is de ETA van de container. Mocht deze (behoorlijk) verschillen dan wordt de klant geïnformeerd. ETD is ook belangrijk maar deze informatie staat al vast en is dan ook alleen voor controle.

Portbase is een systeem waarmee de douane ook werkt waarmee statussen van aangiftes worden beheerd. Voor zover ik heb begrepen is dat het proces globaal gezien als volgt loopt: De klant komt met een nieuwe aangifte, voornamelijk over mail. Deze wordt verwerkt in een nieuwe aangifte in het systeem. Deze aangifte wordt met de bijbehorende certificaten opgestuurd naar de NVWA en de douane. De douane verwerkt een automatische controle, dan wel een fysieke controle. Mocht de container door de controle heen komen, dan wordt de aangifte van de container door de douane op de volgende status gezet waarmee in principe alleen nog maar de logistieke verplaatsing van de container moet worden geregeld.

Tijdens dit proces moet ene T1 document worden gemaakt. Een document T1 wordt opgemaakt voor extern communautair douanevervoer. Dus voor het vervoer van goederen van de ene plaats naar de andere plaats in het douanegebied van de EU, waarbij de goederen onder douanetoezicht zijn gesteld. Dit houdt in dat de goederen vergezeld gaan van een douanedocument, het zgn. T1. Hier moet handmatig enkele details worden ingevuld vanuit Portbase. Vraag is of dit ook te automatiseren.

De gemiddelde tijd per dossier is 45 minuten.

Interview Planner – Container Terminal Twente (CTT)

Held on November 12, 2015

Summary

During this interview information about barge transportation is retrieved. The interviewee works at the planning department.

For barge transportations, they use AIS data websites for current locations and terminal websites to confirm the status of the vehicle. During the trip, the ETA of the vehicle is retrieved via AIS data websites. Important is that locks and tides have big influence on the travel speed. CTT sees the advantages of the platform to increase customer satisfaction by showing them via the portal their shipments. A smaller advantage is the automatic retrieval of the ETA and status since they call regularly with the shipper.

Interview

Dit gesprek heeft plaatsgevonden op het kantoor van CTT in Hengelo. Doel van het gesprek was om voor de modaliteiten binnenvaart en spoor meer informatie over de informatiebehoeften en processen te winnen. Dossiers worden 1x per dag afgegaan en kost per man ongeveer een half uur.

Binnenvaart

Voor de binnenvaart heeft CTT zowel import als export processen. Beide processen zijn grotendeels hetzelfde. Daarom zal het import proces wat uitgebreider worden besproken dan export.

Voor import geldt dat ze voornamelijk terminal websites in de gaten houden. Hier kunnen ze zien wanneer containers zich op een schip bevinden dat zij verder transporteren landinwaarts. Hoewel ze een grove planning hebben wanneer welke container richting Twente moet, kan dit ook worden aangepast als een schip later of eerder aankomt. In eerste instantie kijken ze naar de status van een schip, wanneer het schip is aangekomen wordt gekeken of de containers zijn uitgeladen. Wanneer dit zo is, dan kunnen de containers worden opgehaald. Dit wordt gedaan op een niet openbaar gedeelte van de ect website, namelijk de ‘Container Status Inland Carrier Import’ pagina. Vanuit daar wordt via Marinetraffic gekeken of het schip daadwerkelijk in de buurt is.

Het vertrek van een schip wordt vastgesteld met de informatie die terminal websites als ect aanbieden of Portbase. Hiermee wordt de status opgehaald van een schip. Dit wordt gecontroleerd via AIS data van Marinetraffic. Gedurende de reis wordt gebruik gemaakt van dezelfde AIS data om de ETA in de gaten te houden.

Sluizen onderweg kunnen een obstakel zijn gedurende de reis die voor flinke vertraging kunnen zorgen. Zeker gedurende hoog en laag water kan dit verschillen. Wanneer het water hoog is en de stroom mee dan kan een schip een stuk sneller varen dan andersom. Wanneer het water laag is, dan kunnen schepen namelijk minder vracht mee en is het drukker op het water, wat resulteert in wachtrijen bij sluizen.

Sluizen die schepen onderweg tegenkomen zijn in Hagestein, Amerongen, Driel, Eefele en Delden, waarbij de laatste twee voornamelijk voor flinke vertraging kunnen zorgen.

Waar zij voornamelijk winst zien in het automatisch track en tracen is het feit dat dit ook gepresenteerd kan worden richting de klant.

Voor de export geldt het eigenlijk andersom, alleen wordt in het dossier opgenomen wanneer een schip is vertrokken en niet van een website wordt afgehaald. Gedurende de vaart wordt voor locatiebepaling en ETA bepaling gebeld met de schipper. Omdat het verschil van eb en vloed zodanig groot is en ect hier geen rekening mee houdt met hun ETA bepaling wordt er vertrouwd op het inzicht van de schipper. Hier wordt verder ook AIS data gebruikt.

Interview Planner Import – HST Groep, Sea- and Airfreight Division

Held on September 23, 2015

Summary

The interviewee works at the import deep sea department. Currently, they use three website types as information source, the terminal websites, carrier and AIS data websites. For the terminal websites it is important to retrieve ETA, ETD and the status of the vehicle. The progress of the shipment over a longer period of time is checked by the carrier websites and the AIS data websites. The AIS data websites show the current position of the vehicle while the carrier shows the planning of the vehicle. The shipment has three different statuses: Pre-booked, Scheduled and Arrived. For the customer it is currently not possible to see whether the shipment has departed. Therefor the addition this extra status in the application is desirable.

Interview

De importafdeling van zeevracht is onderverdeeld in twee types: De binnenkomende boten en de uitgaande boten. Het verschil is dat de binnenkomende boten binnen komen in Rotterdam terwijl de uitgaande boten vertrekken richting Rotterdam maar voorlopig nog op vertreklocatie zijn. De schepen die gecontroleerd worden vallen in een tijdsperiode van 7 dagen vanaf heden.

Voor de boten die binnen komen bij Rotterdam gebruikt de gebruiker een drietal sites die geordend staan in de mate van belangrijkheid:

1. ECT.nl
2. APMT.nl
3. Marinetraffic.com

Elk schip/deeltransport wordt in het Microsoft Dynamics weergegeven. In het overzicht staat de verwachte aankomsttijd (ETA) in Rotterdam. Voor elk schip wordt vergeleken of de ETA in Dynamics actueel is met de ETA in de terminalwebsite ect.nl. Mocht het niet in ect.nl staan, dan wordt apmt.nl geraadpleegd. In het geval dat het daar ook niet weergegeven wordt, dan wordt marinetraffic.com aangeroepen. Gekeken wordt niet alleen naar de ETA maar ook of het schip daadwerkelijk is gearriveerd. Is dit het geval, dan wordt de status van het schip aangepast naar gearriveerd.

Bij websites als ect.nl wordt ook aangegeven wanneer het schip weer is vertrokken (ETD) vanuit de terminal. Dit is belangrijke informatie omdat deze informatie wordt gebruikt om aan te geven aan partners dat de container uitgeladen is en beschikbaar is om op te halen.

Bij de uitgaande boten wordt vooral op gelet of het schip op het juiste moment vertrokken is vanuit de plaats van vertrek. Hier wordt voornamelijk Marinetraffic.com gebruikt voor boten waarbij binnen twee dagen bij een port moet aankomen of zijn vertrokken. Mocht echter blijken dat de port waar het schip heen moet of vertrokken is verder light van 1 port, wat betekent dat de destination of de laatste port call niet overeenkomt met het systeem, dan wordt het reisschema opgezocht van het schip. Daar wordt gekeken naar de port die in het systeem staat en wordt de bijbehorende datum vergeleken met het systeem. Voor schepen die verder dan 2 dagen verwijderd of vertrokken zijn van een port, daar worden automatisch de reisschema’s opgezocht.

De websites die gebruikt worden voor het opzoeken van reisschema's zijn vooral:

- HAPAG
- Hanjin
- Maersk
- Bij uitzondering: UASC

Om aan te geven welke informatie belangrijk is voor de gebruiker bij de import is hieronder een kort overzicht gegeven:

Aankomst terminal	Naam Schip	Vertrekdatum schip	Vertrektijd
ETA	ETD	Plaats bestemming	Aankomst dag
Aankomst datum	Aankomst tijd	Aankomststatus	Vertrek dag
Laatst bezochte port schip	Geografische positie (uitgaand)	Koers schip (uitgaand)	Reisschema (Uitgaand)

Het proces kent een viertal statussen, waarbij het eigenlijk handig is om een vijfde er bij toe te voegen. De drie processen die ze nu in het algemeen gebruiken zijn:

1. Pre Booking
2. Scheduled
3. Arrived

Handig is het toevoegen van een vierde status, namelijk 'departed' zodat klanten kunnen zien of het schip daadwerkelijk vertrokken is. De laatste status is 'Incident'. Dit wordt gebruikt als er iets gebeurd wat niet voorspeld was, bijvoorbeeld dat een container niet geladen is en op een ander schip wordt opgehaald.

Een probleem dat kan voorkomen is dat de port die in het systeem staat niet overeenkomt met in het schema dan wel marinetraffic overzicht. Voorlopig weet men uit ervaring of de genoemde port overeenkomt.

Interview Planner Export – HST Groep, Sea- and Airfreight Division

Held on September 23, 2015

Summary

The interviewee works at the export division for deep sea. For the export division it is important to know whether the vehicle has departed and the container is on board. The core service of the division is the confirmation of the departure. The export division delivers an extra service to their customers with the confirmation of the arrival. The confirmation is retrieved by a Master B/L. Before the retrieval of the Master B/L they check only the status of the vehicle.

Interview

Bij de export afdeling van de zeevracht wordt vooral gekeken of een schip vertrokken is waar een container op zou moeten zitten en wanneer de BL informatie beschikbaar is, gechecked of de container daadwerkelijk op het schip zit. Het verifiëren of het schip vertrokken is met de container is de core service wat ze leveren, het verifiëren van de aankomst van het schip bij de bestemming is meer een extra service wat ze leveren aan de klant.

Voor schepen die vertrekken vanuit Rotterdam wordt voornamelijk gechecked op Master BL of container nummer op Track-trace.com. Hier wordt gechecked of daadwerkelijk het schip vertrokken is of dat het enige vertraging heeft. Dit wordt ook vergeleken met de data in het backoffice systeem. Mocht het niet te vinden zijn op track-trace, dan worden websites van de container terminals aangeroepen. Hierbij worden vooral ect.nl en apmt.nl voor gebruikt. Echter de Master BL informatie verkrijgen ze pas een paar dagen naar het echte vertrek van een schip. Daarvoor kan alleen bekeken worden of het schip daadwerkelijk vertrokken is via marinetraffic.com. Dit heeft echter als nadeel dat er geen confirmatie is of de container ook daadwerkelijk geladen is op het schip.

Een Master BL bestaat vaak uit een 4 cijferige SCAC code + booking nummer. De SCAC code is de Standard Alpha Carrier Code en is een unieke code waarmee transport bedrijven kunnen worden geïdentificeerd.

Voor de aankomende export schepen worden voornamelijk de rederijen voor aangeroepen. Eerst wordt het geprobeerd via track-trace met de Master BL code. Wanneer dit niet lukt wordt de bijbehorende rederij aangesproken. Mocht dit niet lukken, dan wordt Marinetraffic gebruikt, maar ook nu heb je niet de confirmatie of het daadwerkelijk zo is.

Een lijst met rederijen voor containers (FCL):

- Maersk
- Hyundai Merchant Marine
- Cosco
- PIL
- MSL
- CMA-CGM
- Yang Ming
- OOCL
- UASC

- U Line
- Hanjin
- APL
- Hapag-Lloyd
- NYK
- China Shipping
- Zim
- Safmarine

Voor pallets (LCL) de volgende rederijen worden gebruikt:

- CO loader
- FPS
- Cleve
- Eculine

De informatie die voor export van belang is:

Aankomst datum/tijd	Vertrek Datum/Tijd	Schip naam	Container Nummer
BL nummer	ETA	Plaats van bestemming	Plaats van vertrek (vrijwel altijd Rotterdam)

Interview Planner Airfreight – HST Groep, Sea- and Airfreight Division

Held on September 23, 2015

Summary

The interviewee works at the airfreight division and handles import as well as export shipments. The track and trace process for both import as export are almost identical. For each shipment, the airway bill is leading while the flight number is a nice to have. Via the carrier the status of the shipment is retrieved. When this data are not up to date ADS data websites are used to check the status although this only check the status of the flight and not the status of the cargo.

Interview

Ook luchtvracht heeft te maken met import en export. Beiden zijn vrijwel identiek maar de andere kant op. Een vlucht brengt vracht van A naar B met een eventuele tussenstop in C. In het backoffice systeem wordt altijd het laatste vlucht ingevoerd. Dus de vlucht van A naar C wordt niet ingeboekt. Dit betekent niet dat ze dit niet in de gaten houden. Op de Airwaybill staan ook de verschillende vluchten.

Voor elke order de Airway Bill is leidend. De vluchtnummer zijn altijd nice to have. De airwaybill wordt ingevuld bij track-trace.com bij Air cargo. Dit levert vrijwel altijd een resultaat op. Hier staat op wat de status van de order is. Mocht deze verre van up te date zijn, dan wordt via Flightradar gecheckt of de vlucht vertrokken is. Ook hier is dit niet een goede graadmeter aangezien Flightradar niet aangeeft of de lading aan boord is. Via track-trace.com kan dit wel worden gedaan.

In het systeem kent met een drietal statussen:

- Scheduled
- Departed
- Arrived

De informatie die ze gebruiken voor het bepalen of het volgens plan gaat:

Airway bill nummer	Vluchtschema	Status

Appendix B: Identification of Current Track and Trace Processes

This appendix elaborates over the different processes planners use to track and trace their shipments. For each company each of the processes will be shown and explained.

HST Groep

HST Groep has Deep Sea and Air shipments where they use data from the internet to track and trace their shipments. For each modality, the processes are explained. When necessary, the import and export processes are separated.

Deep Sea: Import

The import section of sea freight is divided in two smaller groups of ships which are checked twice a day. The first group consists of ships that are incoming at Rotterdam within seven days while the second group consists of ships that are leave their departure location within seven days. These ships are checked once in the morning and once in the afternoon. All the ships that are in between are not checked, since those have low to no priority. Those ships are split since they use different data sources. Firstly, the incoming ships at Rotterdam are discussed.

Incoming ships to the Netherlands

For each incoming ship, planners retrieve trip specific information via their back office system Microsoft Dynamics. Information used is the ship name and the scheduling information retrieved from the terminal websites. For the import department the ETA and the status in the back office system must be up to date.

For each incoming ship, the ship name is retrieved and filled out in one of the two terminal websites to retrieve more accurate information at the ETA and the status of arrival. The website of the Europe Container Terminals (ECT.nl) is mostly used since it delivers the most accurate data for most ships. If ECT.nl has no ship information, the website of APM Terminal Rotterdam (APMT.nl) is used. If that does not the ship information, the RWG website (rwg.nl) is accessed. In rare occasions, the ship information is not found at the three websites. In this case, the website of Marine Traffic (Marinetraffic.com) is used. The data found are compared with their back office systems. If the data are not up to date, the data are updated.

ECT.nl stores ETD information which HST use for communication with their partners. When a ship has departed, it means that the container is off the ship and ready to be picked up. If a ship has not departed yet, it is possible that the container is still on the ship and the transportation company that handles the rest of the transport is waiting for the container. Therefor the status of the ship changes to ‘arrived’ when a ship has departed. Figure 48 shows the step-by-step process.

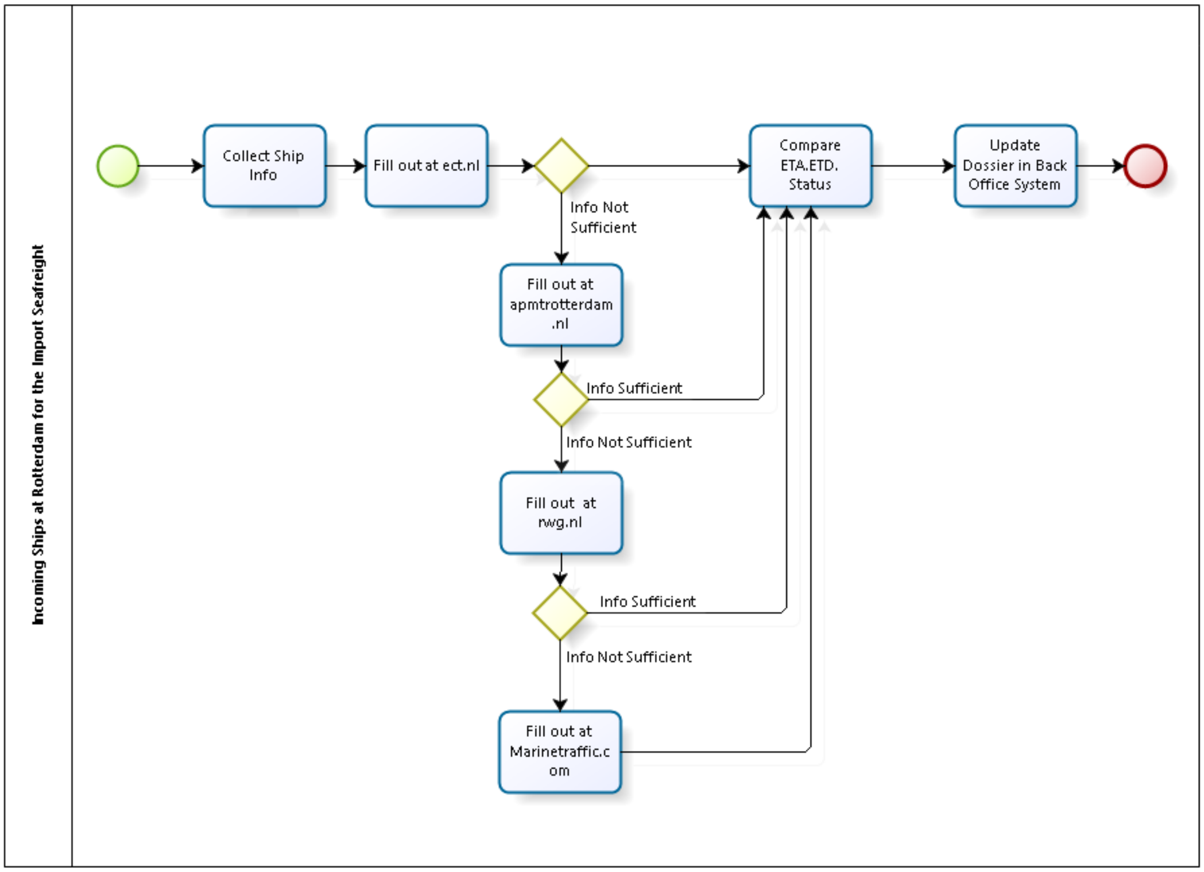

Figure 48: Step-for-step process at HST Groep for incoming ships for import seafreight

On all the different websites different information is stored but not all information is used. Two planners are asked what data they use and need to decide whether or not to update the ETA. Combining the results, the information needed for import over Sea for incoming ships at Rotterdam is shown in table 37.

Table 37: Information used at HST Groep by planners at import over sea department

Arrival Terminal	Name Ship
Date and Time of departure of the ship at Rotterdam	ETA
ETD	Arrival Destination
Day/Date/Time of arrival	Arrival Status
Day/Date of departure	

Outgoing ships leaving their departure location headed to the Netherlands

The process for outgoing ships that are headed to Rotterdam is different. For incoming ships the data at Marinetraffic.com were not accurate enough to determine whether or not a ships has arrived at Rotterdam so they rely more on terminal websites. For outgoing ships this is the other way around.

Planners check whether the ship has departed from the departure location or whether the ETA for the ship to arrive at the departure location is still up to date.

Terminal websites in Asia and other parts of the world are not as fast as here in the Netherlands updating terminal information. Another problem is that it is too time consuming to search for each ship the specific terminal website. This means that for the outgoing ships they mostly rely on Marinetraffic.com or if this information is not sufficient, the trip schedules of the ships are used. If the next port in the back office system has a more than one port call difference, then trip schedules are used to check the ETA at the port that is in the back office system as well. Secondly, if the port call in the back office system is over more than two days, then the trip schedule is automatically used.

Trip schedules are retrieved from the website of the transportation companies. Mostly the schedules are found at Hapag-Lloyd.com, Hanjin-Shipping.com or Maersk.com. If not found at these websites, they find the schedules at smaller transportation company websites as UASC. The step for step process for outgoing ships is shown in figure 49.

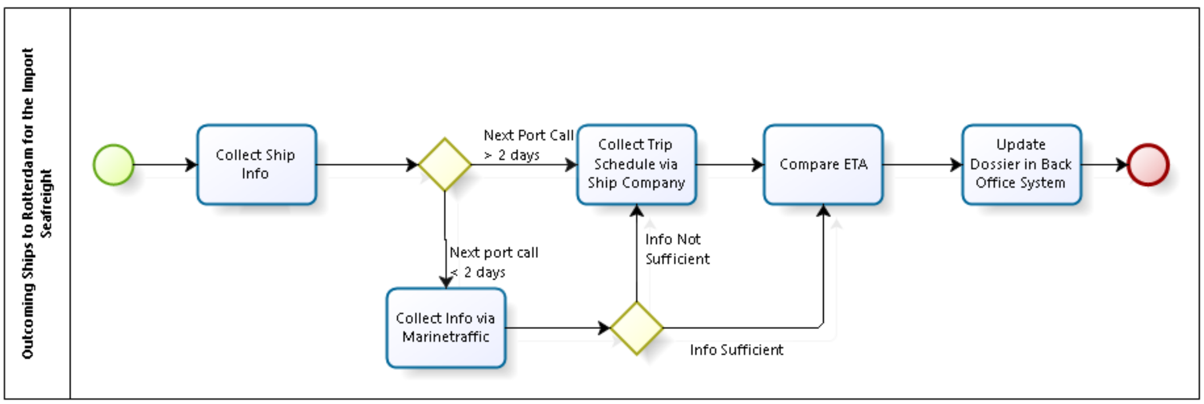

Figure 49: Step-for-step process for outgoing ships for import seafreight

For the outgoing ships, the two planners are asked again what data they use and need to decide whether or not to update the ETA or status of the trip. Combining the results, the information needed for import over Sea for outgoing is shown in table 38.

Table 38: Information needed for Import over sea for outgoing ships

Ship name	Trip schedule
Geographical position	Course Ship
Last Visited Port	Next Port Call
ETA	

Deep Sea: Export

Import focused on ships with a destination at Rotterdam while export focuses on ships that departure from Rotterdam. The export departure is responsible for checking the departure of a ship and inform the customer if the container is indeed on the right ship. As an additional service, they also monitor whether the ship has arrived at its destination. This means that that the export department has two processes, one to determine the departure of the ship (with the container) and one to determine the arrival of the ship. They check each ship within 7 days of arrival or departure twice a day.

Outgoing ships from Rotterdam

The Export Department determines for each trip if the ship has departed or it has any delay. The tracking of the outgoing ships is done in different ways. The most reliable and accurate way is to track and trace the ship via the Master Bill of Lading (B/L). Although it is not certain whether a container is loaded, it is desired to track the departure of the ship. If the Master B/L is not received, the data of the container terminals as ect.nl of apmt.nl are used. If the terminal information is not sufficient, marinetrafic.com is used to track the location of the ship. When the Master B/L is received, this document is leading. The step for step process is shown in figure 50.

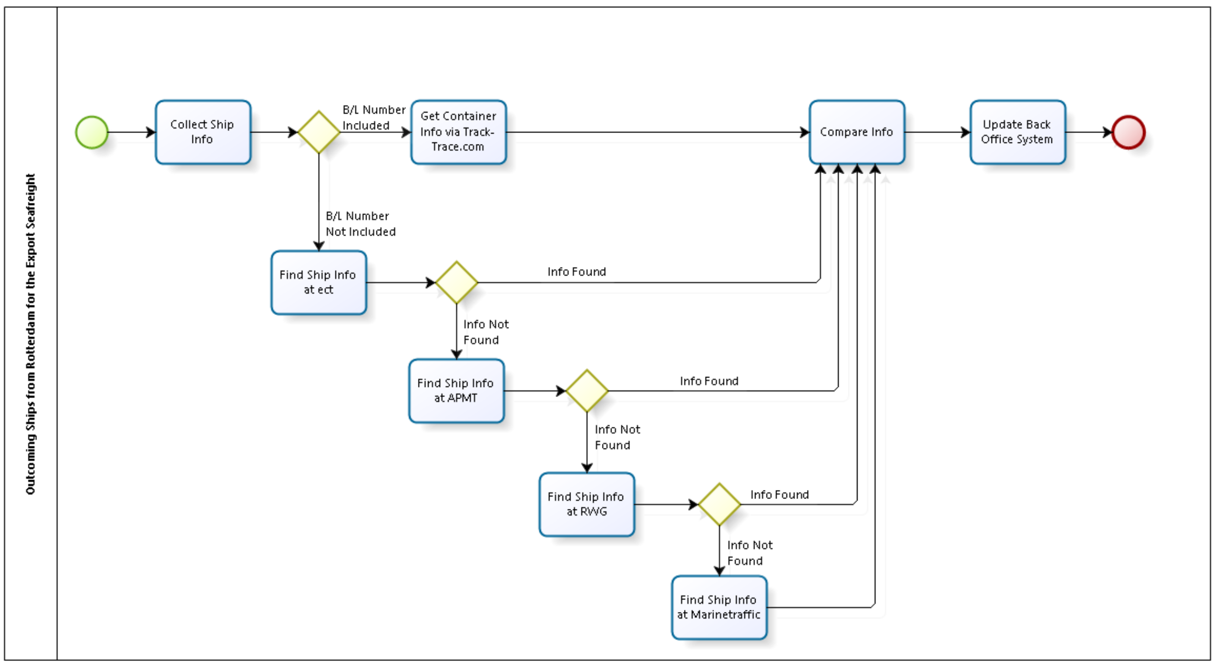

Figure 50: Step-for-step process for outgoing ships from Rotterdam for export seafreight

Two planners of the export department are asked again what data they use and need to decide whether or not to update the status of the trip. Combining the results, the information needed for Export over Sea for outgoing ships is shown in table 39.

Table 39: Information needed for the export department for outgoing ships

Scheduled Arrival Data/Time	Scheduled Departure Data/Time
Ship Name	Container number
Master B/L Number	Scheduled ETD
ETD	Destination
Departure location	Geographical Information
Course Ship	

Incoming ships at a foreign port

An additional service HST SAF offers is tracing the ship until it has arrived at its destination. For this process the Master B/L is again important. At the moment the ship is close to its destination HST has the Master B/L in its position. Via Track-trace.com the Master B/L is used to find the container- and trip information. The track-trace website routes the user to the right port website. This is the port of the destination. The ship information is connected to the Master B/L and made available by the port. Not all ports are even fast updating their system. If that is the case Marinetraffic.com is used to check its current port, destination and its geographical location but this information is less accurate and therefor used less. The process is shown in figure 51.

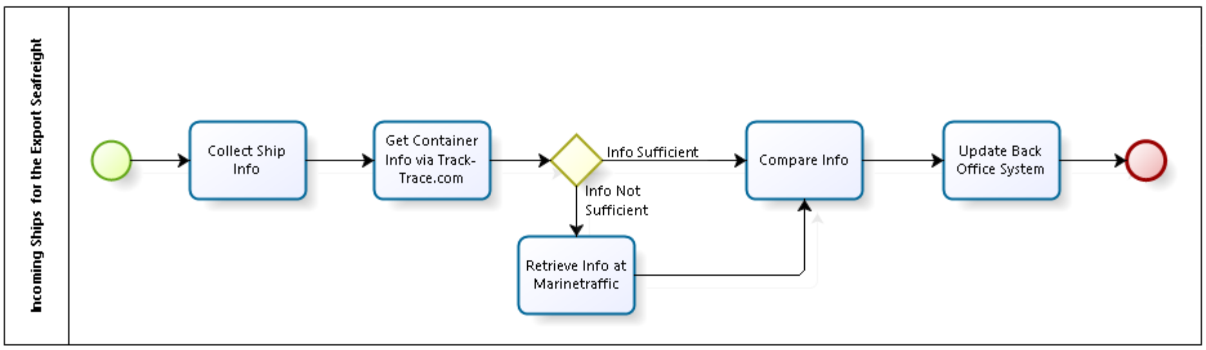

Figure 51: Step-for-step process for arriving ships for export seafreight

Again, two planners of the export department are asked what data they use and need to decide whether or not to update the status of the trip. Combining the results, the information needed for incoming ships is shown in table 40.

Table 40: Information needed for the export department for incoming ships:

Scheduled Arrival Data/Time	Scheduled Departure Data/Time
Ship Name	Container number
Master B/L Number	Scheduled ETA
ETA	Destination
Departure location	Geographical Information
Course Ship	

Air: Import and Export

Planners use the AWB number to track the shipment via track-trace.com. Track-trace.com routes the user to the airline company identifying the AWB number. The airline gives regular status updates, although some airline update their status slower than others. This means that information can be out of date. If the information is not up to date and the flight should have been departed, flighradar24.com is used to track whether the flight has taken off but this website is seldom used. Based on the information of the airline, the status is changed. The process is shown in figure 52.

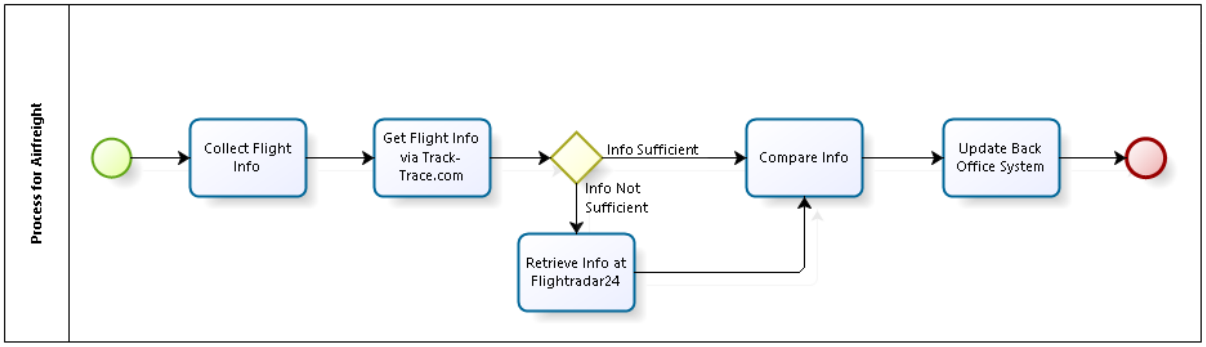

Figure 52: Step-for-step process for airfreight

Two planners of the airfreight department are asked what data they use and need to decide whether or not to update the status of the trip. Combining the results, the information needed is shown in table 41.

Table 41: Information needed by airfreight planners

AWB Number	Status
List of Flightnumbers	Arrival location(s)
Departure location(s)	Arrival Date/Time(s)
Departure Date/Time(s)	

Planners update the status in the system if the airplane departed from its departure location and has arrived at his final destination. The Airfreight department has three different statuses:

1. Scheduled
2. Departed
3. Arrived

Unlike by sea transport it is common that freight travels over different airplanes. This means it can go from Amsterdam to Brazil via Spain with different airplanes. The back office system cannot five detailed information about where the freight is. HST communicates the freight has departed and change the status to 'Arrived' when the freight is ready to pick up at its final destination. In between more detailed information can be given to customers to give them more insight where there freight is.

Neele-Vat

Neele-Vat has both Deep Sea and Air shipments. During the interviews only Deep Sea shipments are taken into account.

Deep Sea: Import

The Import department at Neele-Vat is responsible for all incoming orders at Rotterdam. The process starts with an incoming orders and ends with the settlement of the container in the Rotterdam port. The track-tracing process of a ship is important to check whether a ship has arrived and a container is on land. This triggers the declaration process of the container at the Customs division. Legally, this cannot be done before the container is off the ship. The process differs than the processes at HST. HST has two separate processes to check departed ships and arriving ships, Neele Vat uses onl0 yone process.

The moment that an order is made final and the Import receives the B/L document, a dossier is made for each order. In this dossier includes the details of the order as well as the arrival information of the ship at Rotterdam and the container information. This also starts the track-tracing process. For Neele-Vat the container number is the most important variable in the tracing process because it happens more than once that a container switch ships without updating the Neele-Vat. The information at the shipping company is most up to date about the ship the container is at. Once a week the container is checked.

If an order is a within a week of the estimated arrival, the process switches. De frequency of updating the status of the container goes to once a day and the used data sources switches towards the Rotterdam terminal websites ect.nl and apmtrotterdam.nl. The terminal websites gives status information about a boat, the ETA of the boat and the status information of the container.

In an ideal situation the dossier is automatically updates based on the information that starts with the container number and that the system sends a warning when a container has great difference in ETA of

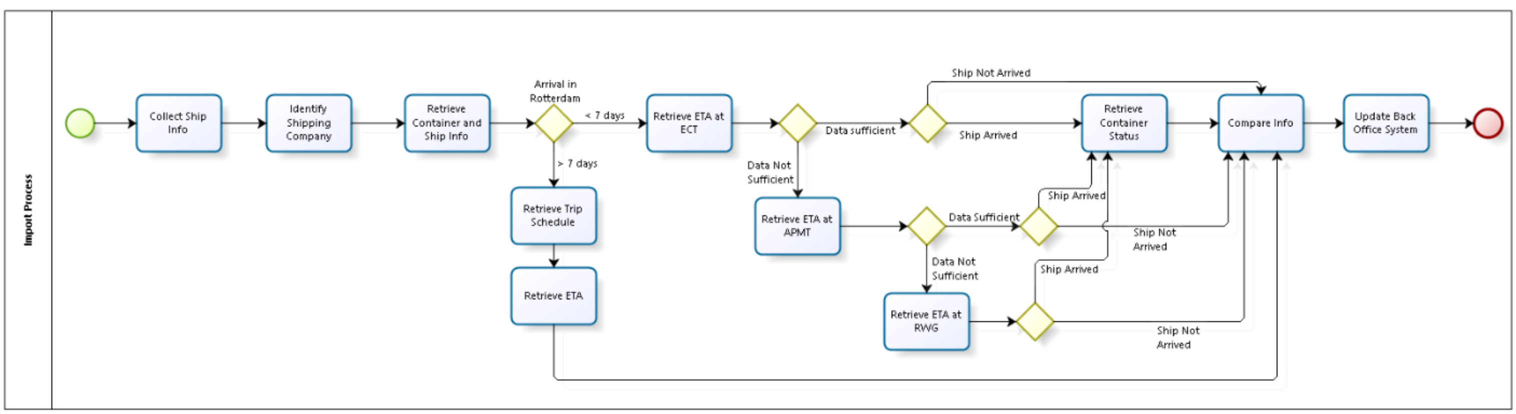

Figure 53: Import Process at Neele-Vat

when a container has arrived and is off the ship. Figure 53 shows the process.

During the interview it was asked what data they use and need to decide whether or not to update the dossier. The information needed is shown in table 42.

Table 42: The data needed for the import process

Container Number	B/L number
Ship Name	ETA
ETD	Ship Status
Ship Schedule	Container Status

Deep Sea: Export

The Export department handles everything for export orders, orders that leave the port of Rotterdam and go to a foreign location. The department handles the customs, planning and booking of a container and ship. This process ends when the ship has left the port, the Master B/L is received and the bill has been paid by the customer. For the planning it is a nice feature when automatically ship departure schedules can be found at shipping company websites to easily find the right departure to meet the conditions of the customer.

Later in the process when the order is final and the container and ship is booked the dossier must be updated periodically, once a day, to update the status of the ship. Important details are the closing time of the ship. This information can be found at the terminal websites. The closing time is the latest time a container can be loaded on board. If a ship arrives a day later in Rotterdam, this automatically means that the closing time changes as well. If the closing time exceeded, it is important to know whether a ship has departed and if the container is on board. The master B/L confirms this a few days after departure but before that the combination of the departure of the ship and whether the container is not in Rotterdam anymore is sufficient. Figure 54 shows the process.

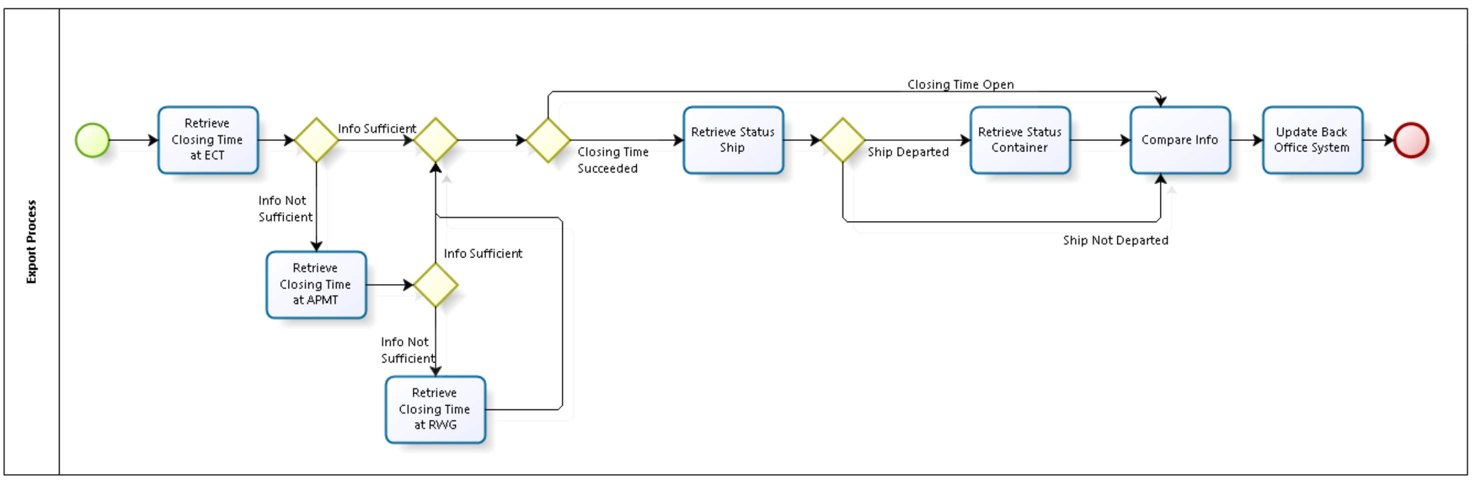

Figure 54: Export process for the periodically process at Neele-Vat

During the interview it was asked what data they use and need to decide whether or not to update the dossier. The information needed is shown in table 43.

Table 43: Data needed for the export process at Neele-Vat

Ship Name	Container Number
Ship Status	Container Status
Ship Closing Time for Boarding Goods	

Container Terminal Twente (CTT)

CTT's main business is the transportation of goods from Rotterdam to their inland terminal at Hengelo or from Hengelo to Rotterdam. This transportation is with barge.

Barge: Import

The import orders are orders that are transported from Rotterdam back to Hengelo. Before a ship leaves with containers to Hengelo, the deep sea ships are checked every day for the status of the container. This creates the possibility to change containers to different ships when a containers is too late or early. Firstly, the status of the ship is checked and when the ship has arrived in Rotterdam, they switch to the status of the container. From the terminal websites they can check whether a container is ready to be picked up.

For the tracking process it is relevant when the ship towards Hengelo is departed. This is checked via the terminal websites and then verified via the AIS data available at Marinetraffic. During the trip, the same AIS data source is used to constantly check the current ETA. The import process is shown in Figure 55.

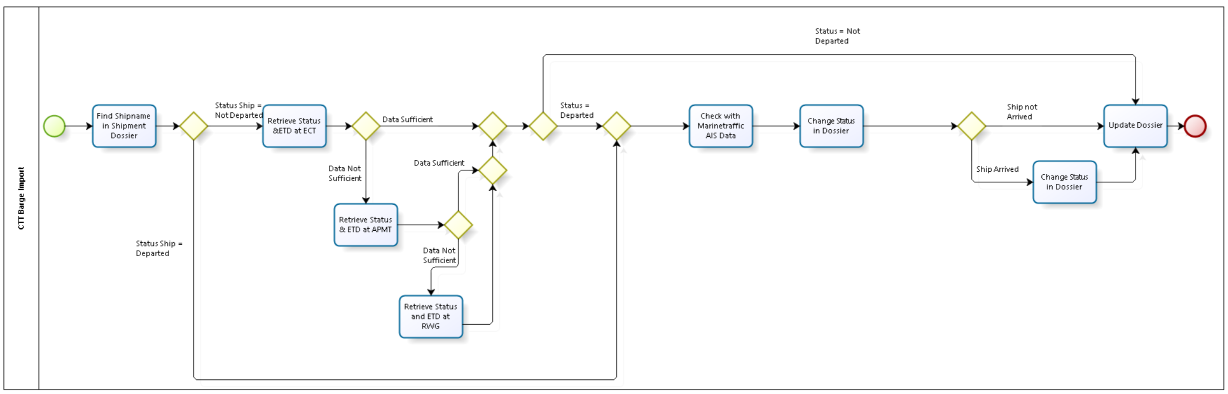

Figure 55: Import process at CTT for barge

The information that is used from the order dossier is needed and updated is shown in table 44.

Table 44: Information needed for barge import process at CTT

Shipname	Status	ETA
From	To	

Barge: Export

For the export orders from Hengelo to Rotterdam, planners have internal information about when a ship has departed. This means that they can not find this information at any information on the website. Their own information is most accurate. When they are notified that a ship has departed, the dossier is manually updated. During the trip the AIS and terminal data are used to track the ship and check the ETA of the ship. However, they notice that these websites are often not that up to date. The ships that are traveling between Rotterdam and Hengelo are often possessed by CTT self. It is therefore easy to communicate and call with the shipper to hear the ETA predictions directly. The export process is shown in Figure 56.

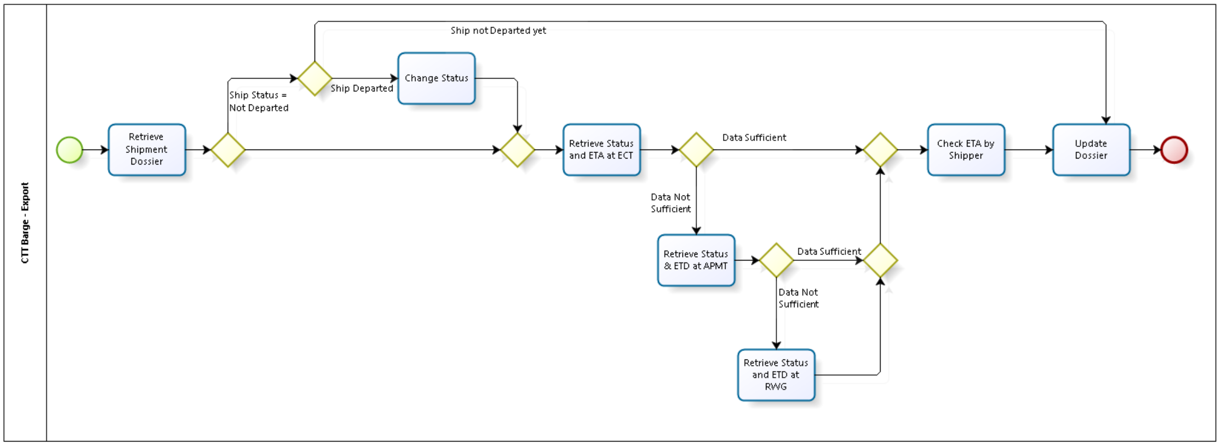

Figure 56: The export process for barge at CTT

The data used are shown in table 45.

Table 45: Data used for the Export process for barge at CTT

Shipname	Status	ETA
From	To	

Locks and high and low tide

The duration of a ship to Rotterdam or Hengelo can fluctuate much when the circumstances are different. Two of the variables that have most influence on the duration of a trip are the locks and the tides.

Inlands waters use locks to overcome the different water levels between different waterways. This means however that is not possible for ships to constantly sail through locks. When it is busy, ships have to wait at locks which can lead to huge delays. This is also influenced by high or low tide. .When the water level is low, ships can have less weight which means that more ships are needed to transport the same amount of containers. This results in a busier water way where the chance is higher that ships have to wait at locks. On the other hand the tides also determines how the water flows. When the water streams inlands, the ships sailing to Hengelo are much faster than the ships sailing to Rotterdam. This is the opposite when the waters streams to the sea.

Appendix C: Architectures

Current Architecture

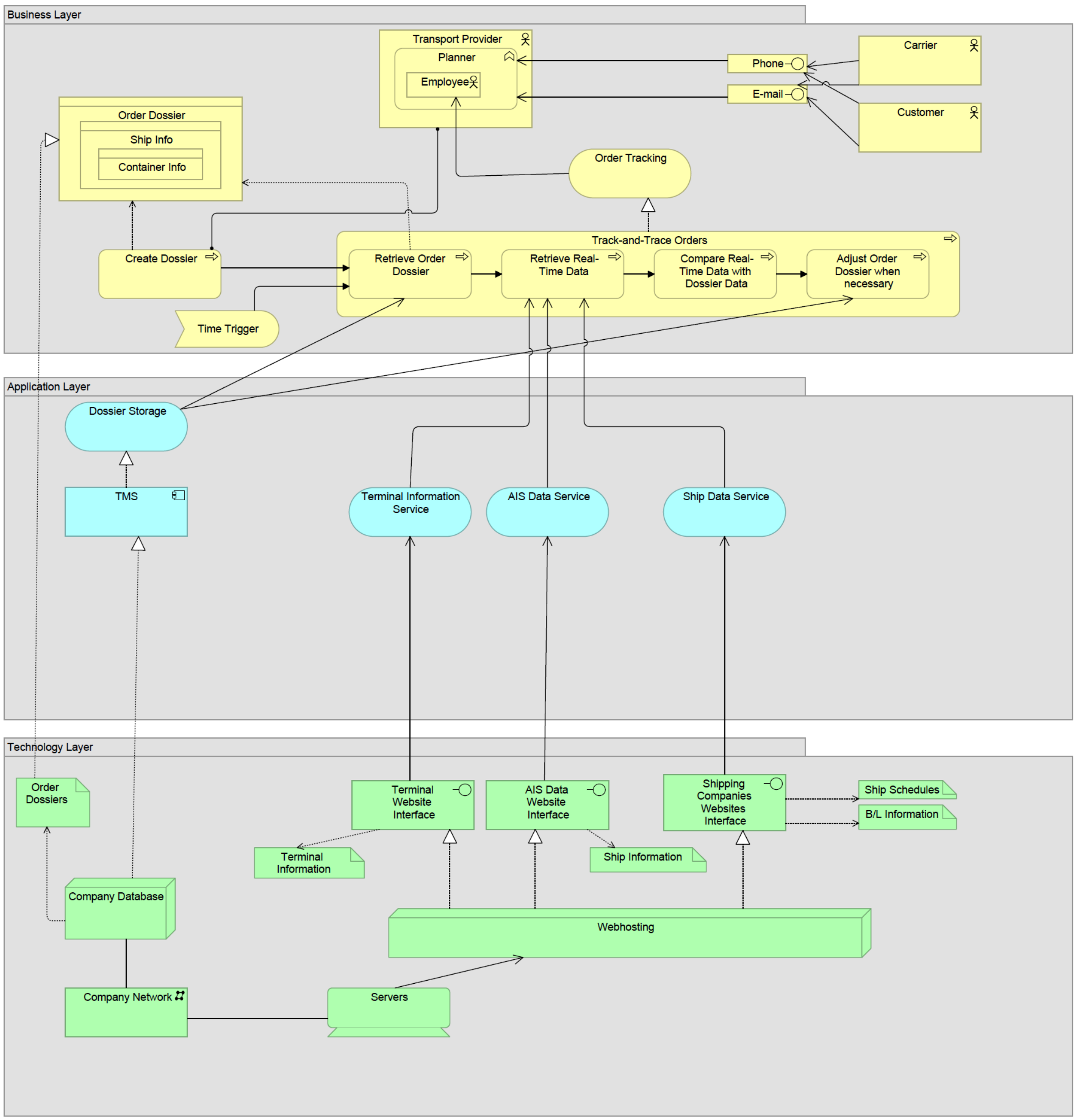

Figure 57: Current architecture of the logistics provider in the supply chian

Target Architecture

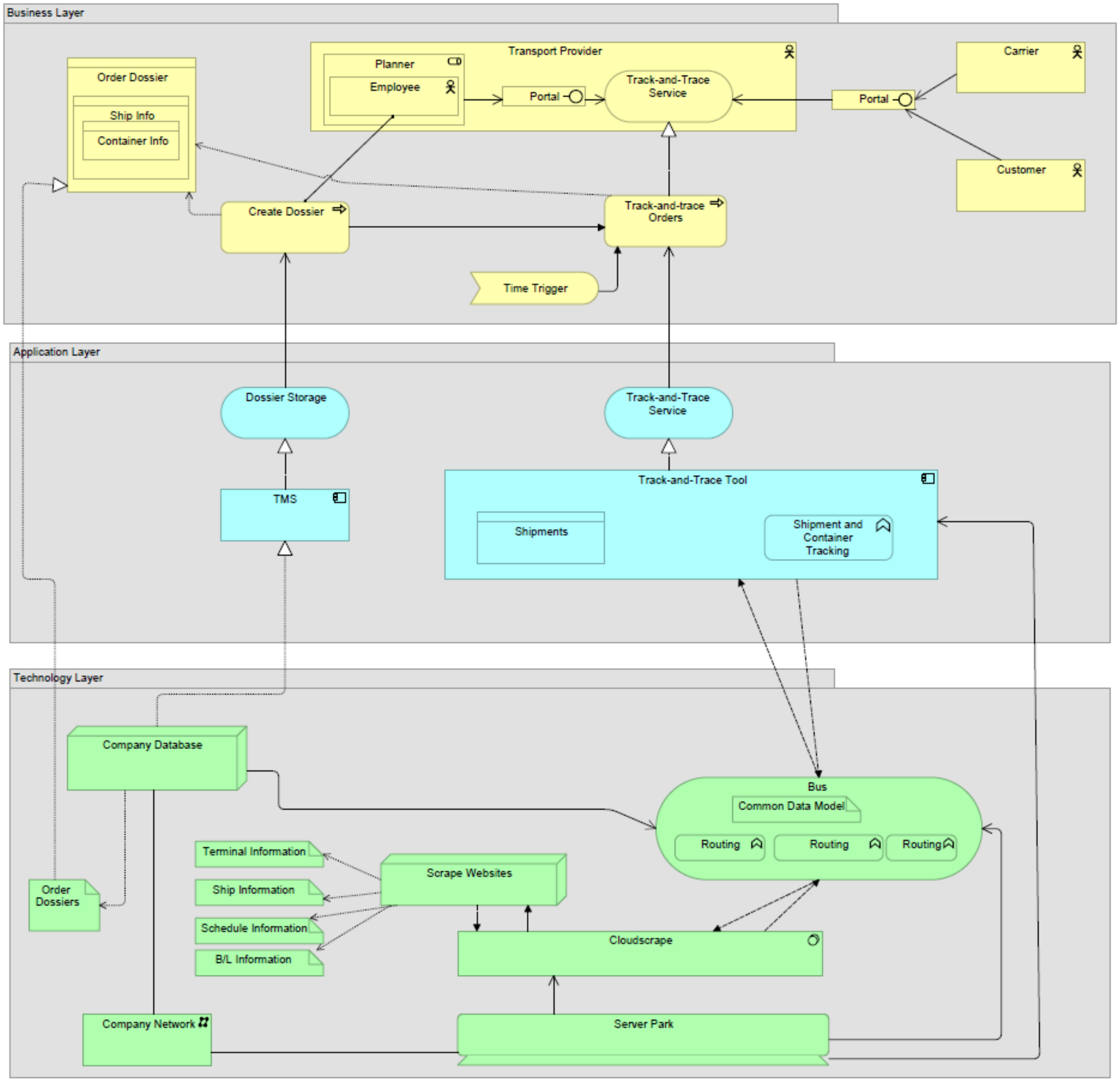

Figure 58: Target architecture of the logistics provider in the supply chain

Gap Analysis

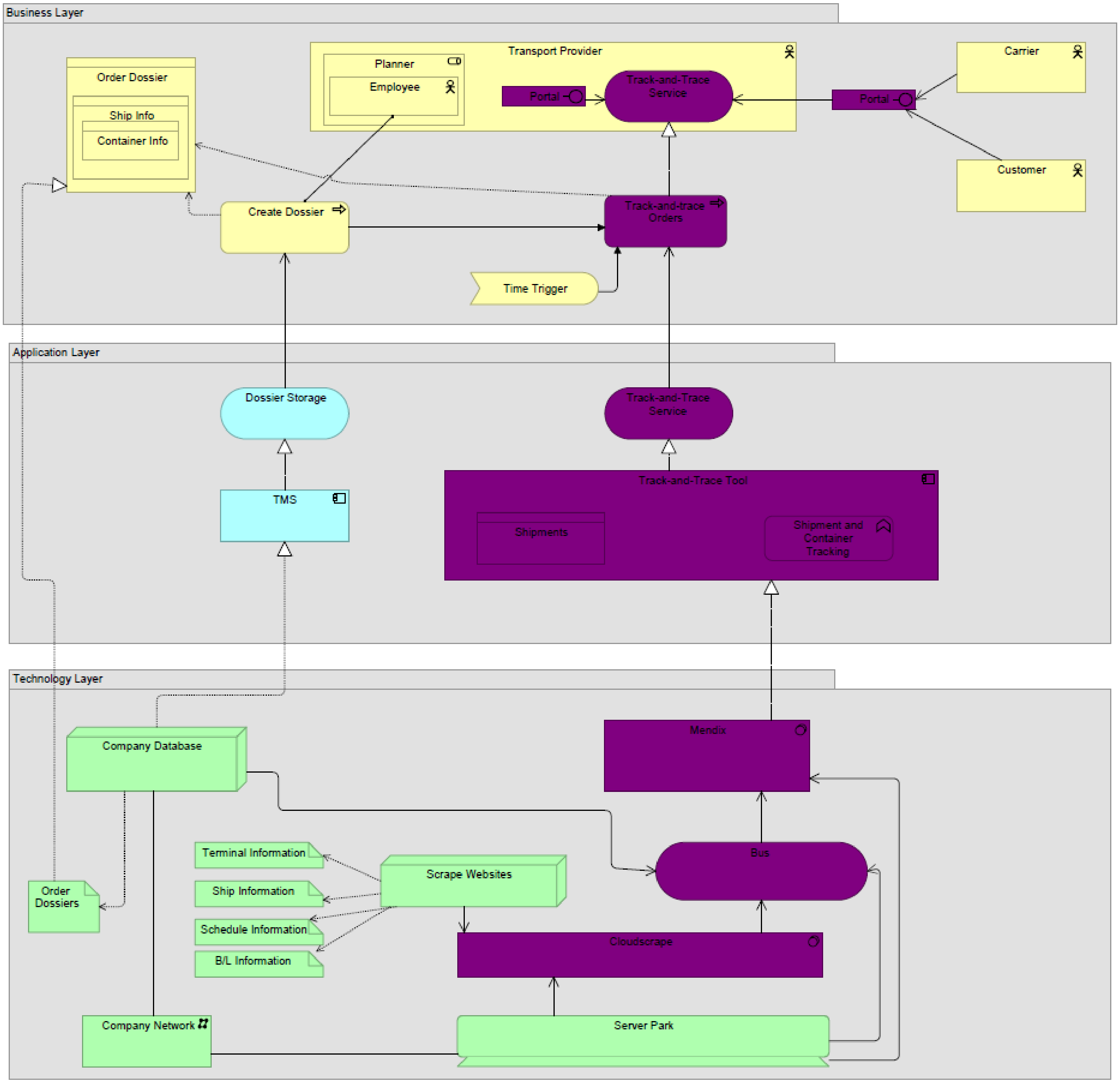

Figure 59: Gap analysis of the current and future architecture

Appendix D: Screenshots of the prototype

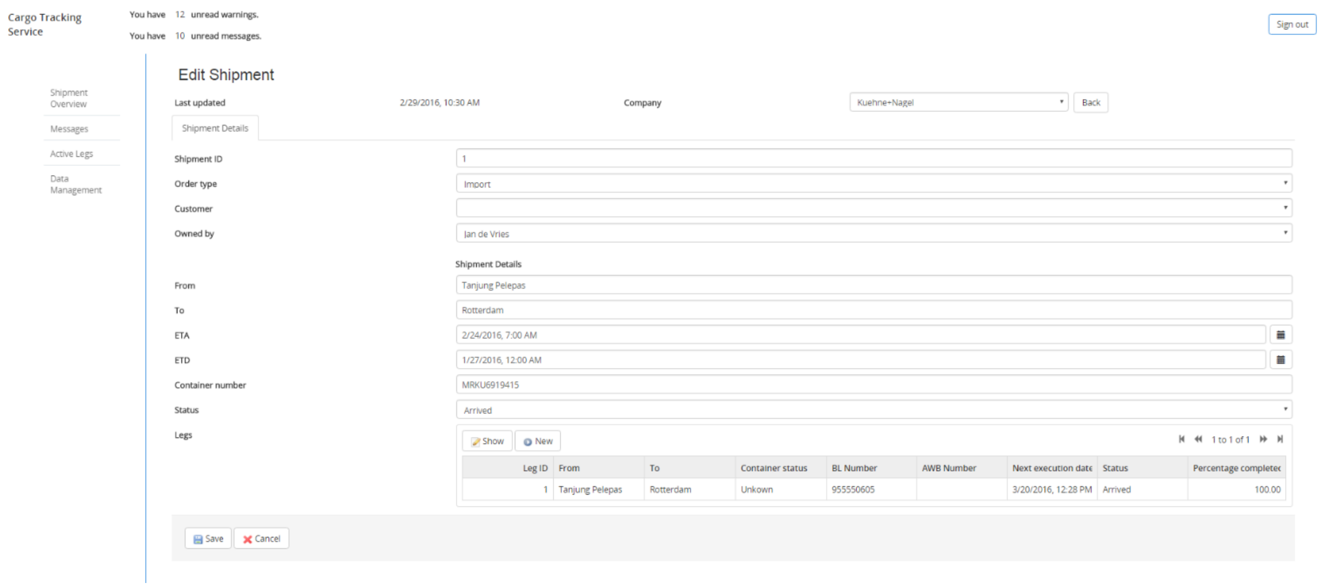

Figure 60: Shipment overview

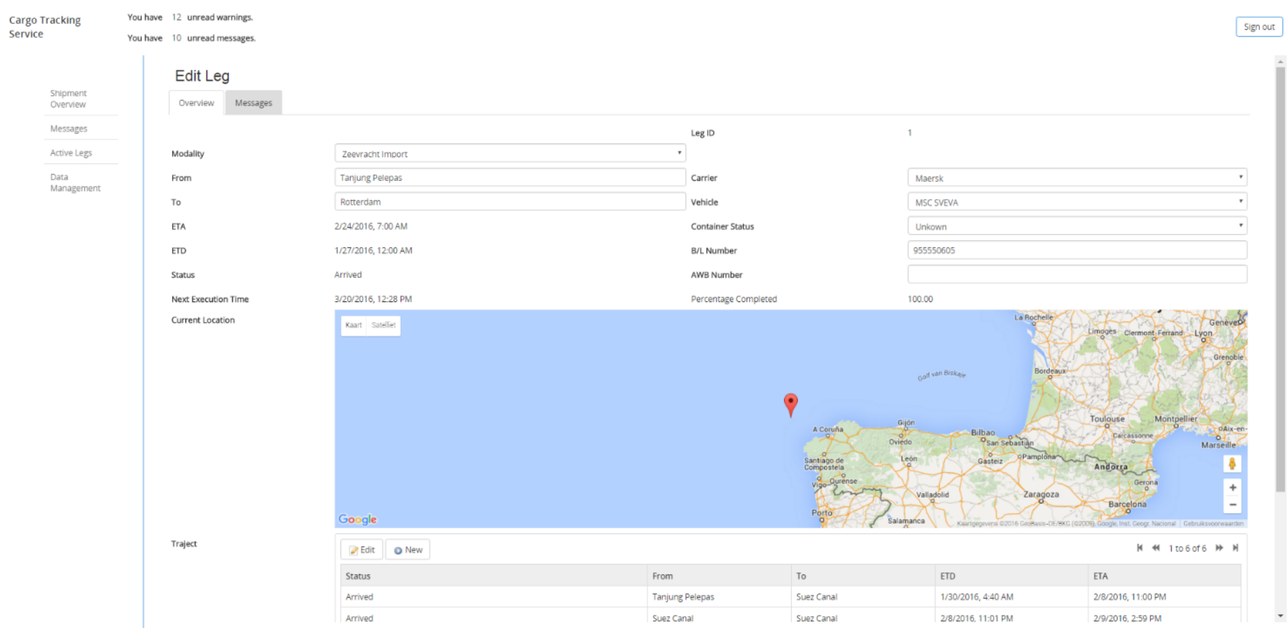

Figure 61: Leg specific Information

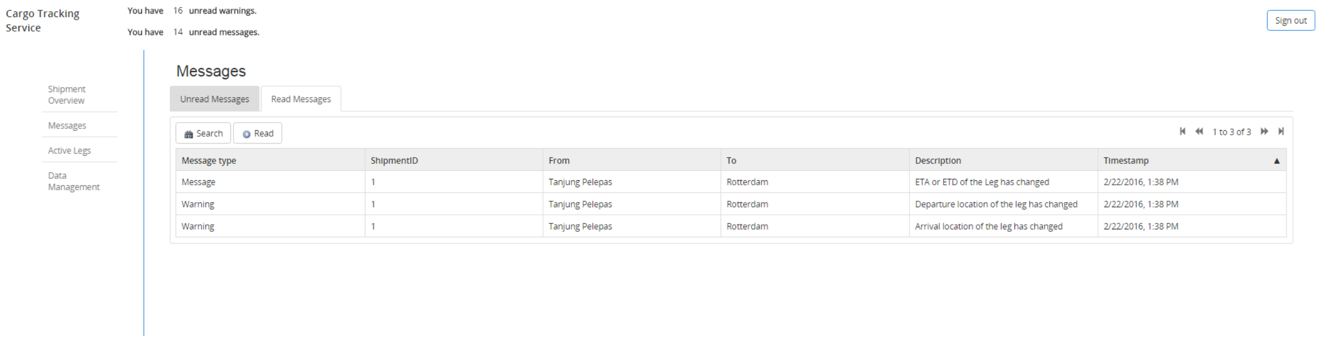

Figure 62: Message overview

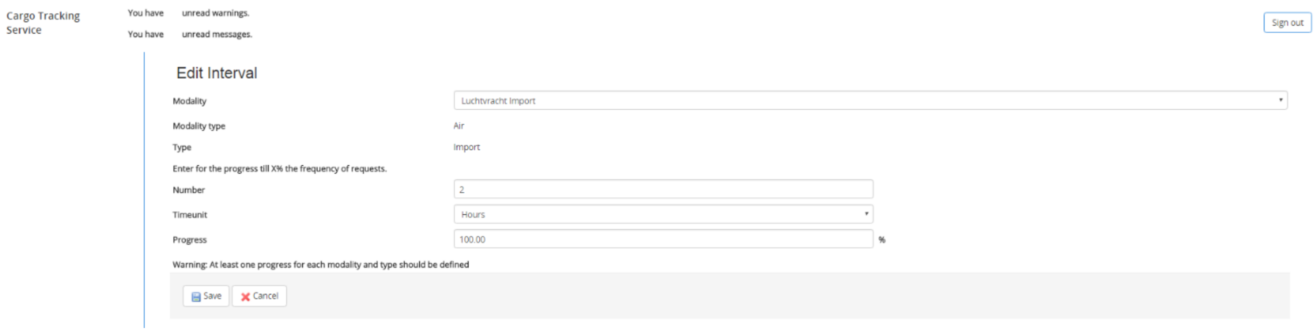

Figure 63: Settings for updating shipments

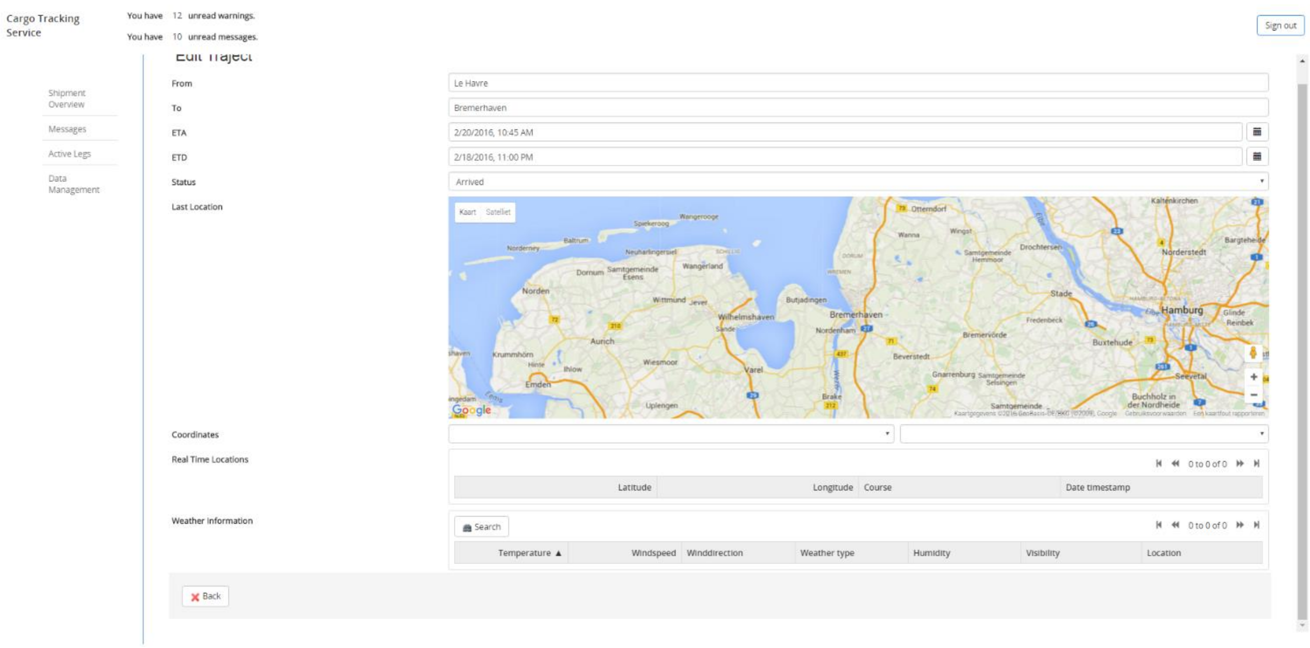

Figure 64: Traject information with weather information

Appendix E: Validation and Evaluation Interview Set-up and Interviews

The questions asked during the semi-structured interviews are listed. The interviews were conducted in Dutch. They are translated into English in this thesis.

At the start of the interview, the setup of the research and the research problem is explained. Asked is whether they still know what is discussed in our previous meeting. Then the prototype is demonstrated by showing the user interface and its functionalities.

After this, the experts were asked questions.

1. Is the information complete in the dossier, and therefor useful to use the prototype instead of manually searching? Is all information present to make good argued decisions? (Corresponds to the Common Data Model)
2. Do you miss certain data representations? (Corresponds to Prototype)
3. Do you miss certain application functions? (Corresponds to Prototype)
4. Do you estimate that the time per dossier will increase, reduce or stays the same by using this prototype? (Corresponds to Prototype)
5. Do you estimate that the ability to execute synchromodal modality shifts will increase, reduce or stays the same by using this prototype? (Corresponds to Prototype)

After these questions, general feedback questions were asked in order to ask about general aspects relating to the common data model and prototype that might be relevant:

6. What are the advantages of this prototype?
7. What are the disadvantages of this prototype?
8. What are the risks of this prototype?
9. Is the prototype an improvement of the current situation? If yes, why?

The application and technology overview of the prototype

Validation with Manager Sea and Airfreight Division and Planner Seafreight - HST Groep, Sea- and Airfreight Division

Interview held at February 28, 2016

During the interview the prototype is demonstrated to Michel and Sjoerd by showing the user interface and its functionalities. After an extensive demonstration, the questions as earlier defined were asked. From here on the results and the feedback are discussed.

The first feedback is that it is, based on the demo, unclear what currently the process is. Questions as 'How is a shipment created' and 'How can I connect websites to the Shipment' are asked. Explained is that the shipments must be automatically created when this is used in practice by connecting the platform with the back-office systems. The websites are automatically connected to the shipments based on the modality defined by the customer. Both agree that there is no business case for this platform when there is no automatic input to the platform.

Based on the presented information, both agreeing that the information is present in the application, although they have preferences for other views of this information. Currently they check the progress of their shipments not per shipment but per vehicle. They would like to have a screen that shows the progress of their current active vehicles instead of legs. This because of that multiple legs can be connected to vehicles and that their focus is on the vehicles within these legs. Therefor it is cleaner if there is a separate view for the active vehicles.

Within HST Groep a shipment is not assigned to one planner but to multiple planners. The platform should be able to do this. Other improvements is that HST Groep would like to set criteria and preferences per customer instead of the currently set criteria. This will include an extra personalization function. Several KPI's must be defined so a dashboard can be generated with performances of the carriers, ships etcetera.

Both Michel and Sjoerd came with extra improvement suggestions. They see potential in the platform when messages can be generated with information that is used to update their own orders in their back-office systems. This will help them to react faster on events and gives them the ability to make more efficient modality shifts since the next carrier can be instructed with more reliable and up to date data. Another suggestions is that customs checks can be integrated as well as widening the scope of the platform with connecting the preliminary stage of an order.

Michel estimates that with the concept with the processes feedback and suggestions can realize a saving of 700 fte yearly. With the processed feedback and suggestions, they are very interested to test this in practice.

Validation with Site Manager Schiphol and Manager Innovation – Kuehne-Nagel

Interview held at March 7, 2016

This conversation is held via Skype. During the presentation of the prototype both are very enthusiastic about the possibilities that the platform offers. Currently planners are checking their flights at the airport Schiphol once or twice a day. They see great possibilities in that flights are continuously updated in the background while the planner works on its other daily tasks.

Ronald mentions that the platform offers a lot of information but that they are currently interested in only expected arrival times of incoming flights. Kuehne-Nagel has currently no synchromodal vision for the platform. They expect positive effects on each of the planner's planning when the information keeps presented in an unambiguously way.

During the interview multiple improvements are suggested. A flight can transport multiple packages. An overview of the statuses of the shipments/packages are less interesting since they give not a good overview of the situation. For each package, they are interested in the ETA and position of the flight. Based on this information they decide to take action. They suggested a portal or dashboard with the statuses and ETA of all incoming flights. This will give a quick overview of the current situation of the flights.

Based on the demo and the interview, Kuehne-Nagel are interested in testing the platform in practice. The test is successful when the platform delivers better, faster and more accurate information that when the data must be retrieved manually and when the users will automatically use the platform instead of manually retrieving the information.