An Architecture and Common Data Model for Open Data-Based Cargo-Tracking in Synchromodal Logistics

Wouter Bol Raap[1,2(✉)], Maria-Eugenia Iacob[1], Marten van Sinderen[1], and Sebastian Piest[2]

[1] University of Twente, P.O. Box 217 7500AE Enschede, The Netherlands
{m.e.iacob,m.j.vansinderen}@utwente.nl

[2] CAPE Groep, Transportcentrum 14, 7547RW Enschede, The Netherlands
{w.bolraap,s.piest}@capegroep.nl

Abstract. In logistics, questions as "Where is my container?" and "When does my container arrive?" can often not be answered with sufficient precision, which restricts the ability of logistics service providers to be efficient. Since logistics is complex and often involves multiple transportation modes and carriers, improving efficiency and saving costs in the supply chain requires communication between the different parties and the usage of real-time data is critical. Currently, logistics service providers (LSPs) use real-time data to a very limited extent, mainly for tracking the progress of a specific part of a given shipment. This data is retrieved manually from a number of websites and sharing with other actors is not even considered. This leads to lack of end-to end visibility and delays in planning. This research proposes an architecture and a common data model for an integration platform that allows the automated collection of real time container tracking data enabling LSPs to plan more efficient. Currently, there is no common data model available that contains all the information required and enables LSPs to track their shipments real-time. The common data model is designed via a bottom-up approach using results of interviews, observations at different logistics service providers, analyses of open data on websites, and serves the information needs of the business processes involving such data. The model is also validated against industry standards. Based on the proposed architecture a prototype was built that is tested in real operating conditions with a fourth party logistics company.

Keywords: Synchromodal logistics · Common data model · Logistics · Integration platform · Web scraping · Open data

1 Introduction

Fourth party logistics companies (4PLs) ideally manage the whole supply chain of their customers, and, as such, they function as sole interface between them and a complex network of providers of logistic services over different modalities: water (sea and barge), air, road and rail [1]. Such a role comes with significant complexities related to the

C. Debruyne et al. (Eds.): OTM 2016 Conferences, LNCS 10033, pp. 327–343, 2016.
DOI: 10.1007/978-3-319-48472-3_19

efficient planning and monitoring of freight movements, while ensuring satisfactory service levels. For achieving this kind of operational excellence, real time tracking data is critical. A concept that expresses accurately this way of working has emerged recently (as generalization of intermodal logistics) under the name of synchromodal logistics. The concept of synchromodality positions the flexible usage of the most efficient mode of transport (with respect to several given criteria, such as, costs, CO2 emissions, service levels, etc.) at all times during a shipment as central management principle [25]. With intermodal transport, the planning of the transport from A to B has fixed times and locations to move the freight from one mode to another [2]. When delay occurs and the container missed the next mode change, the whole planning of the container must be rescheduled, costing time and money. Ideally, the planner wants to plan the container and arrange the transport as late as possible in the process to increase the efficiency of the planning as well as to reduce costs. This requires real-time information that is regularly updated with the latest data. According to [3, 26, 27], synchromodality should make this possible, thus enabling the real-time switching between the transport modes while optimizing multiple criteria, such as costs, sustainability [24], and service levels. Ideally this should be possible in well integrated supply chains with high information transparency. However, in the logistics practice such integration is very rarely achieved due to reasons, such as, dynamicity of the business network (partners in the chain join and leave often), heterogeneity of chain partners (which come from different cultures/countries, have different interests and different degrees of digitization, data they offer differs in quality, real-timeliness and format), low willingness to share information, and unpredictability of operational processes' execution. Thus, currently, most LSPs retrieve and update this data manually from public websites (where more reliable/up-to-date data is offered by third parties, such as port authorities, governments, etc.), which is time-consuming, error prone, and increases the overhead [6].

In 2013, the Synchromodal IT project was started, with the aim to provide a unified platform to integrate various stakeholders in the logistics domain and manage the synchromodal planning process. The added value the Synchromodal IT platform brings is the ability to provide essential information for process optimization that LSPs either could not acquire on their own or the expense of doing so would not justify the potential benefits [28]. As part of this project, the main contribution of this research is an architecture for the automated retrieval of real-time tracking data and a common data model (CDM) for logistics that includes planned and actual information about orders, statuses and disruptions to increase the ability of synchromodal planning of shipments. A common data model is a standardized data model definition for a particular application domain and fosters the transfer of data between data sources and back-end systems. Oude Weernink stresses the importance of a common data model [7] in system integration. Since in our case each tracking data source and back-end system (e.g. planning systems) has its own data format, protocol and encoding, retrieving and using tracking data would be impossible without a common data model. Existing data models currently used in logistics are not sufficient and cannot be used as the common data model for tracking shipments real-time. A CDM will be designed that fits the platform.

The common data model is designed via a bottom-up approach using results of interviews, observations at different logistics service providers, analyses of open data on websites, and serves the information needs of the business processes involving such data. This approach does not rule out an ontology-based data integration approach [30] at a later stage in order to have a more formal content explication and allow easier querying of content. The model is also validated against industry standards. Based on the proposed architecture a prototype was built that is currently tested in real operating conditions with a fourth party logistics company.

Thus, the core novel contribution of this paper is a platform architecture and common data model that makes possible the nearly real-time and synchromodal tracking of shipments, without the need for logistic chain integration, and based on the collection and analysis of open streaming data. The platform (and its tested prototype) is capable to handle open data coming from different sources and of different formats due to the designed CDM, which forms its very core.

The overall research methodology followed in this study is the design science methodology [29]. For the design of the common data model we follow the domain model development methodology of Böhmer et al. [11] to design the common data model containing all the business objects in the selected domain and their connections. This methodology consists of four phases (Problem and Requirements Definition, Analysis, Design, and Maintenance) which can be iteratively applied to build a model that specifies all real and virtual entities that are relevant in a selected domain and structures them by the sets of attributes, reusable sub-components and relationships to other objects.

The paper is structured as follows: Sect. 2 presents proposes and motivates the architecture specification of the integration platform. In Sects. 3 and 4 we present the data requirements, and the CDM definition respectively. These sections cover the four steps of Böhmer's method. We demonstrate the architecture and CDM by means of a prototype, the description of which is given in Sect. 5 together with a discussion of the first test results in a real operational environment. Finally, Sect. 6 concludes the paper with an extensive discussion of possible future work directions based on this research.

2 Architecture of the Integration Platform

In this section we motivate the need for an integration platform in the logistics sector. In addition we explain the architecture of such a platform and its impact on the typical enterprise architecture of a 4PL. We also briefly discuss how unstructured data mined from the web can be presented unambiguously using a common data model, which is necessary for dealing with data heterogeneity.

2.1 Motivation for an Integration Platform

An increased ability to be synchromodal can provide multiple benefits to all the parties involved in the supply chain. This could be the delivery of an improved service by more effective logistic flows, reduced operational risk, increased knowledge sharing, reduced stocks, reduced CO2 emissions, and reduced costs [4]. Synchromodal transport gives LSPs the freedom to deploy different transportation modes flexibly, while increasing the efficiency of the physical infrastructure usage [5].

The decision to switch between different modalities may depend on actual circumstances, such as estimated time of arrival, traffic information, instant availability of assets or infrastructure and all other factors that might change the initial requirements. This requires the latest (if possible real-time) logistics information (e.g., transport demands, traffic information, weather information, etc.) to be available for the planner [12]. However, this is often not available to 4PLs and supply chain partners because of a lack of transparency in supply networks [8]. The main reason is a very heterogeneous IT-infrastructure, and lack of 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 [9]. Some of this information is sometimes available on corporate or public websites, and could, in principle, be collected via an integration platform that monitors shipments about their status and whereabouts.

With the real-time aspect of synchromodality, 4PLs need to act fast and be resilient to sudden changes. 4PLs need therefore a reliable decision support system [8]. Thus, an integration platform that serves as a decision support system has the potential to be helpful and successful in the logistics sector [9] and had been characterized by Vivaldini et al. [13] in terms of the following benefits for the LSP:

- it leads to greater service reliability,
- it makes the shipper more dependent on the LSP,
- it favors the integration with the client, and
- it leads to better delivery information, and fleet management and, thus to operational excellence.

Because the party responsible for the synchromodal planning of shipments is the 4PL, the focus in the remainder of this section is on how the current enterprise architecture of the 4PL is impacted by the implementation of the integration platform. The language we use to specify enterprise architecture is ArchiMate [23]. To design the architectures, the method of Iacob et al. [22] is used.

2.2 Current Architecture

The business layer of the current architecture shows how organizations currently carry out the updating process (Fig. 1). 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 orders up to date in the back-office systems and for interacting with the other two actors (i.e., the carrier and the shipper), to get/communicate these shipment updates.

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. On specific times (e.g., daily, at 9 AM) the updating process is initiated. Necessary data is manually searched and retrieved from certain websites, compared to the dossier in the system and when necessary, changed to the more accurate recent values.

In the application layer, four services are identified that support the updating process in the business layer. The first service is the dossier storage. When a new order is created

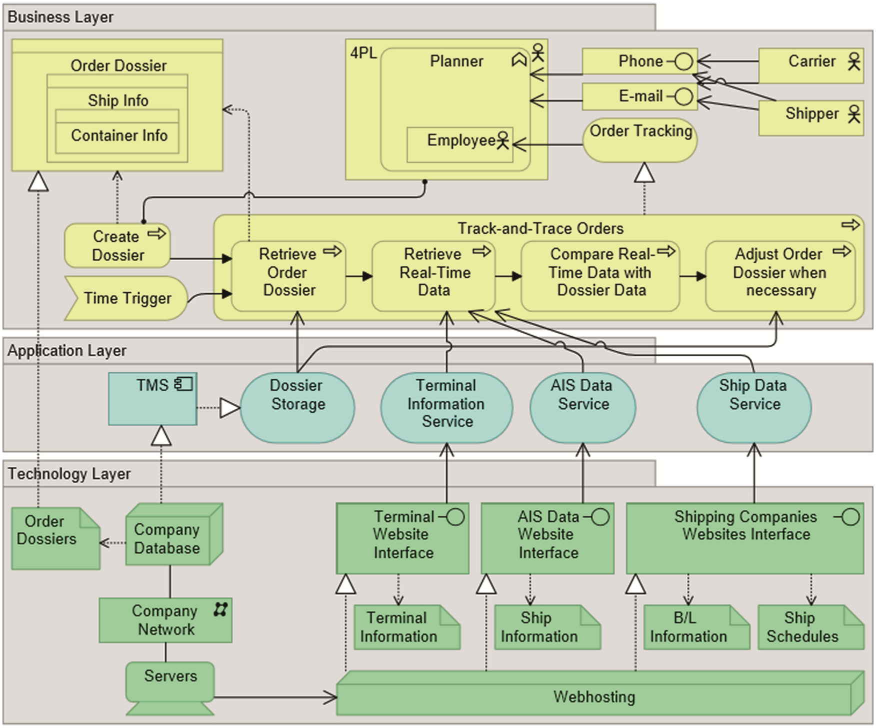

Fig. 1. Current typical 4PL architecture.

a dossier is stored in the Transport Management System (TMS) (which is a common back-office system for any LSP). During the updating process, the dossier is retrieved from the database and updated. The other three services are information services that planners usually use to manually retrieve current shipment information from different websites (e.g., Automatic Identification System (AIS), terminal and ship data).

The TMS runs on top of a database management system (modelled in the technology layer), which stores all the (historical) transaction data of the company. The database runs on internal servers from the logistic company making sure that the data is not openly available for others.

2.3 Target Architecture

The target architecture is shown in Fig. 2. The business layer contains the same three actors. While the planner still has the responsibilities to create, and check the status of order dossiers in the TMS, the planner communicates with a portal for the status of the shipment. The portal is fed with data from the integration platform and 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 tracking information to the actors any time/real-time.

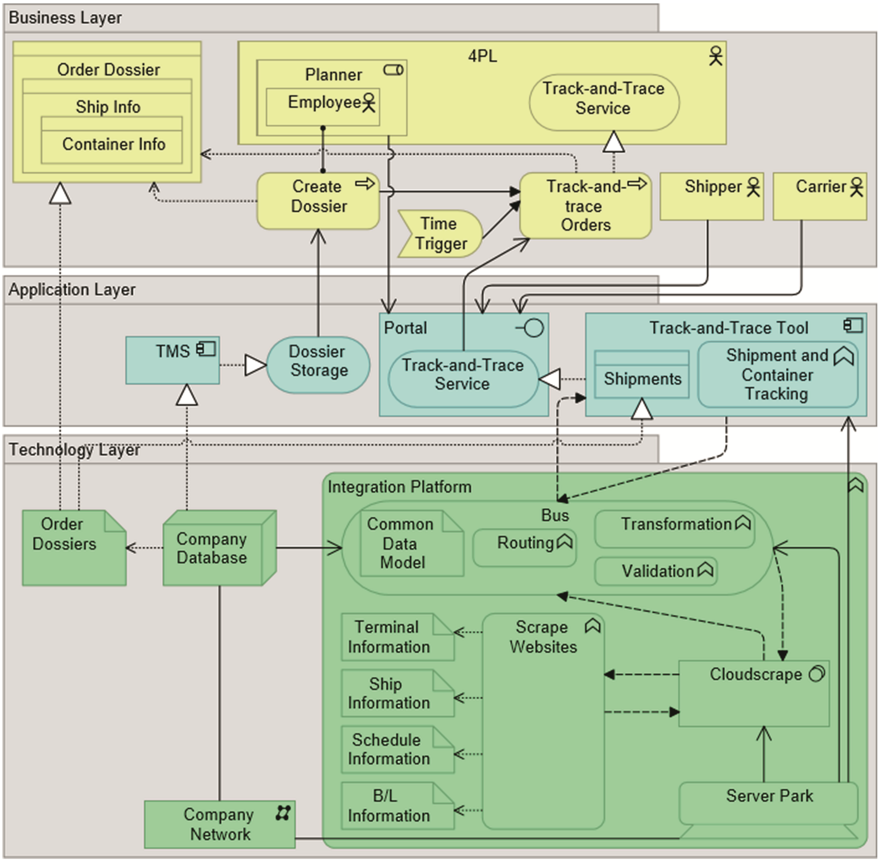

Fig. 2. Target 4PL architecture including the integration platform.

The portal is the interface of the track and trace service. This service updates automatically every 5 min and/or upon request the status of all active shipments, and makes this information available via the portal to all three actors. The service is realized by an application component that is responsible for the automatic retrieval of the information from websites and which implements the business logic to process the retrieved information. The application also makes sure that the information in the portal is shown unambiguously to the user. The other service in application layer (Dossier storage) is inherited from the baseline architecture.

Besides the 4PL's order dossiers database, the technology layer fulfills the function of integration platform. An integration bus manages the communication between the track-and-trace tool and the external information services. More precisely, the integration bus is responsible for the routing, validation and transformation of messages, and is connected to the application in the application layer, the company database and the web scraping tool. From the company database newly created dossiers are retrieved continuously, and send to the track-and-trace tool, which maintains the list of active shipments. The bus communicates with the web scraping tool for the retrieval of information from websites. The web scraping tool scrapes the selected websites that offer the information.

2.4 Dealing with Data Heterogeneity

At the very core of the integration platform is data transformation, which transforms data from a source in such a way that it can be handled by the rest of the integration platform [7, 11]. A common data model facilitates the easy transformation from one format to another and, therefore, fosters the communication (through the integration platform) between back-end systems and data sources [10, 14, 15]. Zaiat et al. [9] state that further specification of the data model is necessary depending on practical needs and applications.

A major challenge for the integration platform is the fact that data offered via websites is presented in various formats, such as structured, unstructured or semi-structured. Unstructured data is data that has no identifiable structure and cannot be stored in rows and columns in a relational database [16]. An example of unstructured data is a document that is archived in a file folder, images and videos. Structured data is described as data that conforms to a specification of a schema [17]. A typical example of structured data is a relational database system. Semi-structured data is often explained as "schema-less or self-describing, terms that indicate that there is no separate description of the type or structure of the data" [17]. Semi-structured data does not require a schema definition. A schema definition is optional, so it is possible to structure data. An example of semi-structured data is XML. In XML data can be encoded directly, while an XML Schema defines the structure of the XML document.

The translation of the information retrieved from websites, and its unambiguous mapping onto a common data model requires two steps.

The first step is to structure the data from the websites by extracting data using web scraping or web data mining [18]. Web scrapers simulate the human exploration of the World Wide Web by either implementing a low-level hypertext transfer protocol, or embedding suitable Web browsers [19]. The output of web scraping, is structured data, for example in a relational database that enables the analysis and comparison of data [20]. Data is presented in a simple format that is easy to process and analyze.

The second step is to translate the structured data obtained as result of web scraping to a common data model format. The CDM format is necessary as intermediary format into a two-step transformation process (Fig. 3). First the message is transformed in the CDM format, which in turn can be routed and sent to all the other connected systems by transforming it the formats they require. In this way discrepancies between message formats of different systems can be solved [21]. These discrepancies can be classified as schematic, semantic or data conflicts [14]. A schematic conflict occurs because of the different ways in which the real world can be structured into objects, leading to different data models. A schema modeler may also be limited by the used technology. 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 the result of the fact that different conceptual schemas can have different semantic meanings for the same real-world object. A data conflict occurs

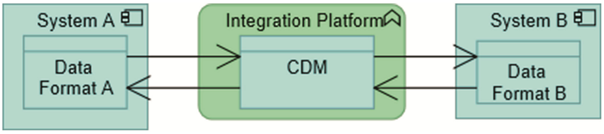

Fig. 3. Data transformations and CDM.

because data is represented in different scales, precisions, or units. For example, the age of a person can be represented by its age or by its birth date.

3 Required Data in the Common Data Model

In Sect. 2 we explained the role of the CDM in the proposed integration platform. However, in order to design this model we first have to understand how 4PLs currently keep track of their shipments. To this end we have observed the planning process at four 4PLs and interviewed some of their planners. Below we report on our findings concerning the identification of required data types that should be covered by the CDM, and concerning the possible sources from where such data can be retrieved.

The frequency with which 4PLs update their shipments varies from once a day to several times a day. However, all 4PLs confirmed the need to acquire the most recent data. One way is to achieve this is by increasing the frequency of updating, which is not feasible in the current setting. A simple calculation shows that if, for example, a 4PL having 1000 running orders (each being shipped from Asia to Europe with an average shipment duration of 4 weeks) which are updated once a day costing 3 min per order, would have to do a total of 28.000 checks costing 84.000 min. This costs the 4PL 125 h/day, which means more than 15 FTEs. By increasing the frequency of manual updates, the costs will increase accordingly.

The daily order update process steps we observed at the 4PLs are comparable and very similar with the reference process shown in Fig. 4.

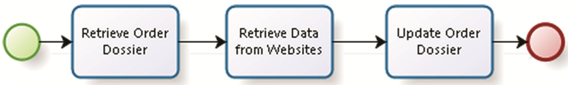

Fig. 4. High level order updating

The first step is the retrieval of the order dossier from the back-end system. In the dossier, the information of the order is stored. Planners retrieve from the dossier the information that websites require as input.

The second step is the retrieval of the latest data from the websites. To retrieve all the information, the planner is required to visit multiple websites. The planner checks information such as the estimated time of arrival (ETA), location of the vehicle or the status of the container.

The last step is updating the dossier in the back-end system with the latest data. Based on the data from the websites, the planner updates the dossier, and decides whether

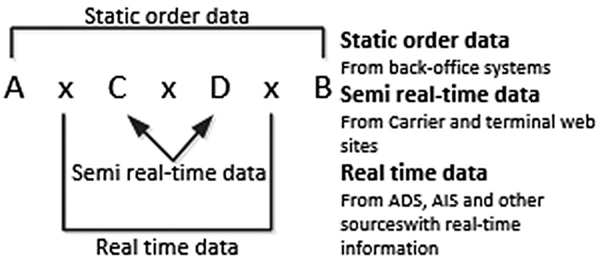

Fig. 5. Categories of data in the CDM.

immediate actions (such as, informing the customer, and the next forwarder, or rescheduling the shipment) are required when delay or other disruptions occur.

It should be noted that planners work with a list of preferred websites where they look regularly for shipment updates. The websites can be classified into four types. Each website contains specific information that cannot be found on other website types. The four website types are:

1. **Carrier** – Holds information about the order based on the order's B/L (Bill of Lading) or AWB (Air WayBill) number as well as vehicle specific data and vehicle schedules. Examples are www.maersk.com for carrier Maersk and www.afkl-cargo.cm for KLM/Air France.
2. **Terminal** – Holds information about the arrival, or the departure of a vehicle at a specific terminal or airport. An example is www.apmtrotterdam.nl, a terminal in the port of Rotterdam.
3. **AIS** – Holds, collects and presents Automatic Identification System (AIS) data of ships. This data is transmitted (as radio signal) by the each ship (e.g., deep sea vessels or barges) and contains data such as ETA, current position, ship identification data, and the next port call. An example is www.marinetraffic.com.
4. **ADS** – Holds, collects and presents ADS (Automatic Dependent Surveillance) data for airplanes. Airplanes are obliged to send data and include information, such as position, and ETA. An example is www.flightaware.com.

The combination of the data types from the order dossiers, and from the different websites types gives a fair overview of the data types needed for being able to execute the updating process in the back-end systems, and which would also be required in the CDM (see Table 1).

Table 1. Required data from websites types and dossier.

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

4 The Common Data Model

As mentioned earlier, to foster communication and data sharing among the supply chain partners, the integration platform requires a CDM. While designing the CDM we followed a three step approach. First, we created a list covering all data types to be included in the CDM. Secondly, entities, and attributes were identified by grouping the above mentioned data types. Finally, the design was completed by defining the relations between entities.

Figure 6 shows a fragment of the common data model specification. The CDM is divided in three areas based on the frequency with which data in those areas change over time. Thus, Static Order Data is data mainly retrieved from back office systems. This data consists of basic order data, such as origin and destination of the shipment, shipper, container identification, trip route, and who is within the 4PL the owner of the dossier. This data rarely changes during the shipment process. This is also the data defining the order, and its trip from the origin A to the destination B (Fig. 5).

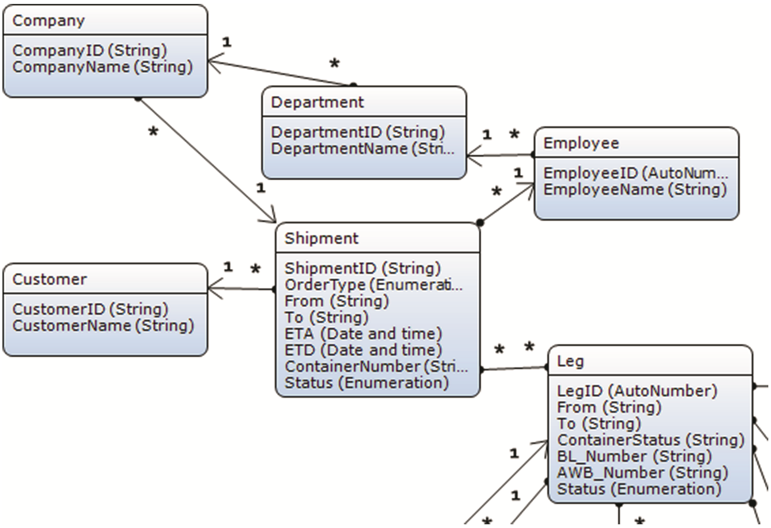

Fig. 6. Fragment of the common data model: static order data

The second type is the Semi Real-Time Data. This includes data that is regularly updated, but changes only occur once or twice a day. This data includes (updated) vehicle schedule, carrier, port calls, and current vehicle carrying the container, and is mainly retrieved from the carrier and terminal website types. With this data, port calls in C and D are also known (Fig. 5). Finally, the third area concerns the Real-Time Data and is mainly retrieved from ADS, and AIS website types. This includes information such as, current location, weather, traffic, and disruptions. This is shown as the real-time, or last known location x of a vehicle between A, C, D and B. Currently most 4PLs use only the first two data types and some 4PLs use some Real-Time Data. This Real-Time Data is most often the location of the vehicle that is tracked. 4PLs have the need to rely more on real-time data in order to make more efficient, and reliable decisions.

The common data model needs to be validated to ensure the completeness of the model. The validation process we carried out included three methods: the Bottom-up Method, the Industry Standards Method, and the Website Method.

With the bottom-up approach we mapped the CDM's against order message definitions originating from the four observed 4PLs, to validate whether the CDM includes all data types required for the updating process. In general, 4PLs use order definitions that include the basic order information to start the updating process. This can be fully mapped onto the CDM.

However, some of the more specific attributes included in the CDM definition (e.g., the owner of the shipment) are not always available in the 4PL message definitions, which means that the CDM is richer than these order messages.

The industry standard approach checks the completeness of the CDM against a sub-set of industry data standards order definitions. This sub-set covers standards that do not include all the required information (e.g., real-time shipment information), but do include the data required to build a dossier and start the updating process. The standards we considered (i.e., IFCSUM, IFTMIN and GS1 Standard Transport Instruction) are a subset of the EANCOM standard. All three of them are EDIFACT standards and define the attributes and structure of an order message. The validation consisted of building mappings between the CDM and the above mentioned standards. IFCSUM and IFTMIN messages can be mapped fairly well on the CDM entities and attributes, although not all the attributes mentioned in the standard definitions are used in practice. GS1 can be fully mapped on the CDM, but the CDM is richer than GS1.

The website approach validates whether the data retrieved from the World Wide Web can be mapped onto the CDM entities. We checked whether Carrier, Terminal, ADS, and AIS websites contain the data we need according to the CDM definition. For Sea, Barge and Air modalities the required data is available. For Sea and Barge however, some weather attributes cannot be retrieved.

Table 2 summarizes the results of the CDM validation as discussed earlier.

Table 2. Applicability of different definitions and websites.

Approach	Name	CDM scope
Bottom-up Approach	4PLs	broader
Industry Standards Approach	IFTMIN	The same
	IFCSUM	The same
	GS1 NL	broader
Website Approach	Websites	broader

5 The Prototype

In this section we present the prototype that has been implemented based on the proposed platform architecture and the CDM. The prototype is used to test the CDM and the functionality of the platform using real data. The positioning of the prototype with respect to the surrounding systems is shown in Fig. 7. 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 the "service" to enable 4PL customers to access the platform. The so-called application layer contains the application logic necessary to process the retrieved data from the websites. This is further detailed in the prototype architecture specification shown in Fig. 8.

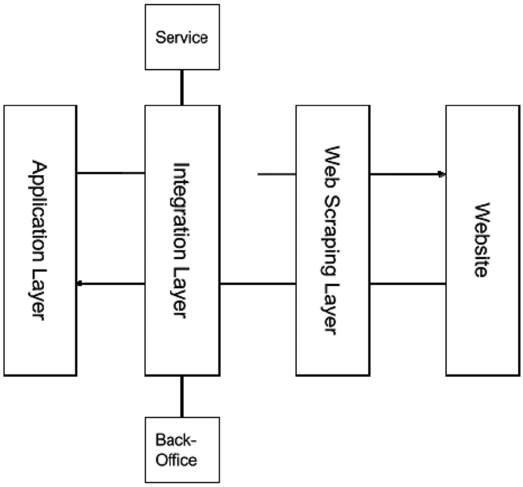

Fig. 7. Positioning of the prototype.

As application development platform we used Mendix, while for the implementation of the integration bus we used the model-driven Integration Platform as a Service eMagiz. The web scraping tool used is in the prototype Dexi.io.

The workflow we implemented for the automated update of orders differs from the manual process. From the interviews we conducted with domain experts it resulted that the identification of the vehicle must be done before the vehicle specific information can be retrieved from external sources. When the vehicle is identified, the rest of the requests will be sent and responses will be collected. Afterwards responses are processed, starting with the information at trip level. Based on the retrieved information the leg will be updated. If necessary, the shipment will be updated as well.

Figure 9 contains a prototype screenshot in which tracking and leg specific information regarding a specific order is shown.

The prototype has been tested during a four week period, using data acquired from one of the collaborating 4PLs at the Schiphol Airport. The scope of the test included the automatic retrieval of ETA information. For the test, shipments using the Air modality are chosen because of the trip duration. This takes several hours on average, and the location of shipments changes rapidly, which makes it the most challenging modality for the test. For Air open data there are three websites types to be considered: Airport, Carrier and ADS websites.

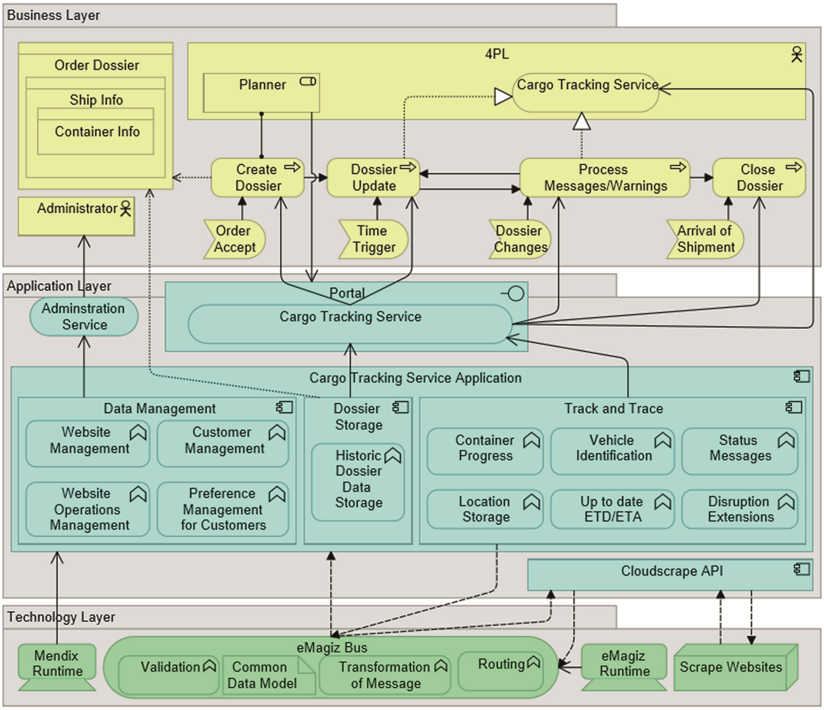

Fig. 8. Prototype architecture.

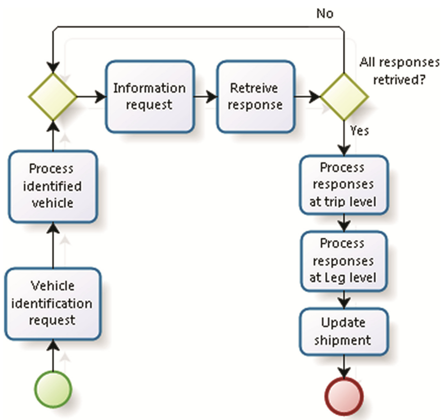

Fig. 9. Implementing order updating workflow.

For Airport and ADS, robots have been built for the retrieval of the data. Carrier websites were not used during the test since they do not provide real-time ETA information. We connected the websites flightaware.com, and Schiphol.nl to the platform.

Schematics conflicts as data representation of flight numbers, and date and time have been resolved by transforming the data to CDM format.

For the test, the robots of the web scraper used a private node at one of the servers of the web scraper. The average duration of a full update of a shipment is 20 s. In practice, with hundreds of shipments, a performance increase is desirable. An increase in the capacity of the private node should lead to performance increase. During the test period the robots were stable. However, they are still sensitive for changes in the website's layout or link changes which causes robots to be unable to find the data. The quick fix for this problem was to manually redesign the robot, which was acceptable considering that this prototype was merely a proof of concept. However, future work must address this issues in a more automated fashion.

Besides these problems, the test period showed that shipments can be tracked automatically over a longer period of time without interference of human planners. The prototype has been tested with real deep sea shipment data with a trip duration of 2–4 weeks, acquired from one of the collaborating 4PLs. The prototype has been also validated through semi-structured interviews. Five domain experts coming from the 4PLs that participated in this study were asked to assess the prototype. They have been selected based on their extensive experience and knowledge of the problem context, and because they could become potential users. One of these experts has worked for four weeks with the prototype during the test period.

The results show that the interviewees see potential in the designed prototype. The interviewees were satisfied with the information made available to the user, but made suggestions regarding its visualization. Some other suggestions referred to the possibility to extend the prototype and the CDM with customs information and make this available from the preliminary stage of an order lifecycle. The test period revealed as well that the user experienced the platform as saving him time, and delivering him and his customers more accurate information than he currently retrieves.

All interviewees estimated that the concept can lead to important savings, as well as to an increased planning efficiency. All the four 4PLs are interested in testing the prototype in real life conditions.

6 Conclusions

In this paper we proposed an architecture, a common data model for logistics that enables 4PLs to be more synchromodal. Available industry standards as IFTMIN and IFCSUM do not contain the information required to track shipments real-time. The common data model designed in this paper currently covers the Deep Sea, Barge and Air modalities. The other modalities Rail and Road have been not included because of limited availability of open data on the web concerning shipment tracking, also because of the limited interest for these modalities manifested by the companies involved in this study. However, for a truly synchromodal CDM, these modalities should also be added.

We have validated the different proposed artefacts both by implementing them in the form of a prototype (which has been tested in an experimental setting), and qualitatively through interviews with domain experts. Furthermore, the prototype will be soon undergo tests in a real-life setting for large order volumes at one of the largest global 4PLs.

The prototype showed that the implemented common data model satisfies the requirements of its intended use. Nevertheless, this prototype should be tested in an operational setting, letting planners use the prototype and quantitatively assess its performance. Besides more testing, future work must also deal the fact that 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.

The scope of the common data model can be much wider to also cover other stages during the order process. Research can be done to identify whether it is interesting to add the planning stage to the common data model, as carriers not only have actual trip information but also planning information available as well. The scope can also be extended with the customs and tax information. Also a refinement of CDM's "disruption" entity might be useful, as disruption are critical for the re-planning of shipments, and thus for synchromodality. Research can be done to specify what types of disruptions are important in practice for each modality, and whether there is data is publicly available.

The platform collects information and stores it in the platform. Another interesting extension to our approach is to develop business analytics functionality on top of the stored data. This data can hold important information and trends that can be derived via data mining. For example, accuracy of planning, performance of carriers, vehicles, suitability of routes etc. could be monitored based on specific KPI's, and may lead to improvements and efficiency gains.

References

1. Runhaar, H.: Efficient pricing in transport. Eur. J. Transp. Infrastruct. Res. **1**(1), 29–44 (2001)
2. Steadieseifi, M., Dellaert, N.P., Nuijten, W., Van Woensel, T., Raoufi, R.: Multimodal freight transportation planning: a literature review. Eur. J. Oper. Res. **233**(1), 1–15 (2014)
3. Van Stijn, E., Hesketh, D., Tan, Y.-H., Klievink, B., Overbeek, S., Heijmann, F.: The data pipeline. In: Proceedings of the Global Trade Facilitation Conference, pp. 1–26 (2011)
4. Egberink, J.: The influence of trust on inter-organizational information sharing in logistics outsourcing relationships. MSc Thesis, University of Twente (2015)
5. van der Burgh, M.: Synchromodal transport for the horticulture industry. MSc Thesis, Erasmus University (2012)
6. Veldhuis, R.: Developing an automated solution for ETA definition concerning long distance shipping. MSc Thesis, University of Twente (2015)
7. Oude Weernink, M.J.P.: Development of an integration platform metamodel. In: Proceedings of the 23rd Twente Student Conference, pp. 1–10 (2015)
8. Toth, M., Liebler, K.M.: Real-time logistics and virtual experiment fields for adaptive supply networks. In: Windt, K. (ed.) Robust Manufacturing Control, pp. 327–340. Springer, Heidelberg (2013)

9. Zaiat, A., Rossetti, R.J.F., Coelho, R.J.S.: Towards an integrated multimodal transportation dashboard. In: Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems, pp. 145–150. IEEE (2014)
10. Singh, P.M., van Sinderen, M.: Interoperability challenges for context-aware logistic services - The case of synchromodal logistics. In: Proceedings of Workshops of the 6th International IFIP Working Conference on Enterprise Interoperability, pp. 1–9. CEUR-WS (2015)
11. Böhmer, M., Schmidt, M., Weißenberg, N.: Seamless interoperability in logistics: narrowing the business-IT gap by logistics business objects. In: ten Hompel, M., et al. (eds.) Cloud Computing for Logistics. Lecture Notes in Logistics, pp. 77–117. Springer, Switzerland (2015)
12. Li, L., Negenborn, R.R., De Schutter, B.: A general framework for modeling intermodal transport networks. In: Proceedings of the 10th IEEE International Conference on Networking, Sensing and Control (ICNSC), pp. 579–585. IEEE (2013)
13. Vivaldini, M., Pires, S.R.I., De Souza, F.B.: Improving logistics services through the technology used in fleet management. J. Inf. Syst. Technol. Manag. **9**(3), 541–562 (2012)
14. Zhang, J.: A formal specification model and its application in multidatabase systems. In: Proceedings of the Conference of the Centre for Advanced Studies on Collaborative Research (CASCON), p. 76. IBM Press (1994)
15. Mangisengi, O., Huber, J., Hawel, C., Essmayr, W.: A framework for supporting interoperability of data warehouse islands using XML. In: Kambayashi, Y., Winiwarter, W., Arikawa, M. (eds.) DaWaK 2001. LNCS, vol. 2114, pp. 328–338. Springer, Heidelberg (2001)
16. Blumberg, R., Atre, S.: The problem with unstructured data. Inf. Manag. (2003)
17. Abiteboul, S., Buneman, P., Suciu, D.: Data on the Web: From Relations to Semi Structured Data and XML. Morgan Kaufmann, San Francisco (2000)
18. Qureshi, P.A.R., Memon, N., Wiil, U.K., Karampelas, P., Sancheze, J.I.N.: Harvesting information from heterogeneous sources. In: Proceedings of the 2011 European Intelligence and Security Informatics Conference, pp. 123–128 (2011)
19. Vasani, K.: Content evocation using web scraping and semantic illustration. IOSR J. Comput. Eng. (IOSR-JCE) **16**(3), 54–60 (2014)
20. Baumgartner, R., Gatterbauer, W., Gottlob, G.: Web data extraction system. In: Liu, L., Tamer Özsu, M. (eds.) Encyclopedia of Database Systems, pp. 3465–3471. Springer, US (2009)
21. Poulovassilis, A., McBrien, P.: A general formal framework for schema transformation. Data Knowl. Eng. **28**(1), 47–71 (1998)
22. Iacob, M.-E., Jonkers, H., Quartel, D., Franken, H., van den Berg, H.: Delivering Enterprise Architecture with TOGAF and ArchiMate. The Open Group presentation (2012)
23. Iacob, M.E., Jonkers, H. Lankhorst, M.M., Proper, E., Quartel, D.A.C.: ArchiMate 2.1 Specification. The Open Group. Van Haren Publishing (2012)
24. Iacob, M.E., Van Sinderen, M.J., Steenwijk, M., Verkroost, P.: Towards a reference architecture for fuel-based carbon management systems in the logistics industry. Inf. Syst. Frontiers **15**(5), 725–745 (2013)
25. Oonk, M.: Smart logistics corridors and the benefits of intelligent transport systems. In: Transport Research Arena (TRA), 5th Conference: Transport Solutions from Research to Deployment, pp. 1–8 (2014)
26. Van Riessen, B., Negenborn, R.R., Dekker, R.: Synchromodal container transportation: an overview of current topics and research opportunities. In: Corman, F., et al. (eds.) ICCL 2015. LNCS, vol. 9335, pp. 386–397. Springer, Switzerland (2015)
27. Lu, M., Borbon-Galvez, Y.: Advanced logistics and supply chain management for intelligent and sustainable transport. In: Proceedings of the 19th ITS World Congress, pp. 1–9 (2012)

28. Dobrkovic, A., Iacob, M.-E., Van Hillegersberg, J., Mes, M.K., Glandrup, M.: Towards an approach for long term AIS-Based prediction of vessel arrival times. In: Zijm, H., et al. (eds.) Logistics and Supply Chain Innovation, pp. 281–294. Springer, Switzerland (2016)
29. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**(3), 45–77 (2007)
30. Noy, N.: Semantic integration: a survey of ontology-based approaches. SIGMOD Rec. **33**(4), 65–70 (2004)