Knowledge Modeling and Incident Analysis for Special Cargo

Vahideh Reshadat, Tess Kolkman, Kalliopi Zervanou, Yingqian Zhang, Alp Akçay, Carlijn Snijder, Ryan McDonnell, Karel Schorer, Casper Wichers, Thomas Koch, Elenna Dugundji, and Eelco de Jong

Abstract The airfreight industry of shipping goods with special handling needs, also known as special cargo, suffers from nontransparent shipping processes, resulting in inefficiency. The LARA project (Lane Analysis and Route Advisor) aims at addressing these limitations and bringing innovation in special cargo route planning so as to improve operational deficiencies and customer services. In this chapter, we discuss the special cargo domain knowledge elicitation and modeling into an ontology. We also present research into cargo incidents, namely, automatic classification of incidents in free-text reports and experiments in detecting significant features associated with specific cargo incident types. Our work mainly addresses two of the main technical priority areas defined by the European Big Data Value (BDV) Strategic Research and Innovation Agenda, namely, the application of data analytics to improve data understanding and providing optimized architectures for analytics of data-at-rest and data-in-motion, the overall goal is to develop technologies contributing to the data value chain in the logistics sector. It addresses the horizontal concerns Data Analytics, Data Processing Architectures, and Data Management of the BDV Reference Model. It also addresses the vertical dimension Big Data Types and Semantics.

Keywords Special cargo · Knowledge acquisition · Ontology · Incident handling · Risk assessment

V. Reshadat (✉) · T. Kolkman · K. Zervanou · Y. Zhang · A. Akçay
Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: v.reshadat@tue.nl

C. Snijder · R. McDonnell · K. Schorer · C. Wichers
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

T. Koch · E. Dugundji
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands

E. de Jong
Validaide BV, Amsterdam, The Netherlands

E. Curry et al. (eds.), *Technologies and Applications for Big Data Value*,
https://doi.org/10.1007/978-3-030-78307-5_23

1 Introduction

This chapter describes ongoing work in the Lane Analysis and Route Advisor (LARA) project which aims at big data analysis and respective knowledge modeling in the logistics sector, namely, in planning shipments with special handling needs known as *special cargo*, or *special freight*, such as cargo consisting of temperature-sensitive pharmaceuticals, live animals, dangerous goods, and perishables, such as lithium batteries, flowers, and food products.

Currently, the execution of such shipments constitutes a complex process that lacks transparency and standardized knowledge resources and relies on the expert knowledge of *freight forwarders*, namely, individuals or companies organizing and planning such shipments. Freight forwarders play a key role in the special cargo industry because they possess expert knowledge on all crucial information for deciding among shipment route options, such as services provided by airlines, cargo restrictions and risks, and transport facilities. For this reason, most logistics operations are handled manually, and there is currently no transparent way of comparing and planning shipment routes, as is the case, for example, with passenger air travel planning.

Route planning for (special) cargo has significant potential for optimization with the application of advanced data analytics and artificial intelligence (AI) methods. However, an added challenge in this application lies in the acquisition and modeling of logistics and cargo knowledge from a variety of available information sources. Currently, standardization and data integration are hard not only due to the data complexity, size, and variation, but also due to cargo service providers attempting to profit from the lack of transparency and information asymmetry. Another challenge relates to processing and classifying cargo information in various types of unstructured, free-text sources, with minimal training or lexical resources. Finally, there are numerous challenges in understanding risks and constraints related to special cargo shipments [12], so as to eventually assess a candidate shipment route. In this chapter, we discuss ongoing work on addressing these challenges.

Our work addresses two of the main technical priority areas defined by the European Big Data Value (BDV) Strategic Research & Innovation Agenda [47], namely, the application of data analytics to improve data understanding and providing optimized architectures for analytics of data-at-rest and data-in-motion, the overall goal being in developing technologies contributing to the data value chain in the logistics sector. With regard to the BDV Reference Model, we address the ''vertical" dimension: *Big Data Types and Semantics*. We also address three 'horizontal' concerns: *Data analytics*, *Data processing architectures* and *Data management*.

This chapter is organized around the three building blocks shown in Fig. 1. In Sect. 2, **Special Cargo Ontology**, we discuss the knowledge elicitation and respective research in modeling cargo knowledge into a standardized form. This work sheds more light on the design and development of a logistics knowledge base and the methodology for eliciting domain information, so as to eventually

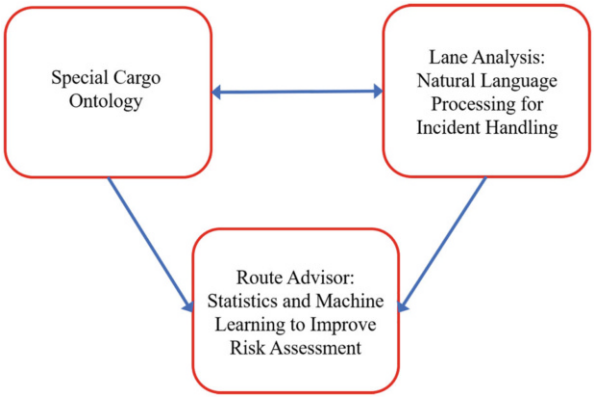

Fig. 1 An overview of the special cargo modeling system

be able to determine routing options. In Sect. 3, **Case Study: Lane Analysis and Route Advisor** we describe a test bed for future application of the knowledge modeling involving the following types of data: structured data, time series data, geospatial data, text data, network data, and metadata. [47] Subsequently, we discuss a novel palate of data analytics approaches to provide a major player in the freight forwarding industry with a set of solutions for several of their organizational issues, using this data. In Sect. 4, **Natural Language Processing for Incident Handling**, NLP and the machine learning algorithm of Random Forests are used to gain new insights on incident classification related to data quality issues in unstructured data. In Sect. 5, **Statistics and Machine to Improve Risk Assessment**, a logistic regression model is used to detect which features most profoundly influence which incident types. With regard to data management, we consider the aspect of data quality affecting the results. The analysis namely has to be considered carefully as data quality issues affect the results.

The chapter accordingly relates to three main cross-sectorial technology enablers of the **Strategic Research, Innovation & Deployment Agenda** for AI, Data, and Robotics, recently released as a joint initiative by the Big Data Value Association, CLAIRE, ELLIS, EurAI and EUrobotics [46]. These cross-sectorial technology enablers are respectively: *Knowledge and Learning* (Sect. 2), *Sensing and Perception* (Sect. 4), and *Reasoning and Decision Making* (Sect. 5). Furthermore, due to the nature of the case study involving incident analysis for special cargo, and thus digital and physical AI working together (Sect. 3), a fourth cross-sectorial technology enabler is inherently addressed: *Action and Interaction*.

2 Special Cargo Ontology

An *ontology* is defined as 'a formal, explicit specification of a shared conceptualization' [35]. One of the main advantages in using ontologies for modeling knowledge lies in allowing a versatile representation of concepts and hierarchical concept relations, properties, and constraints [1]. This also allows machines to make use of the World Wide Web without any interference of humans, as an ontology translates human concepts in machine-readable terms. For our purposes, a special cargo ontology is intended as a knowledge structure that models special cargo services and properties, so as to (1) have an explicit model of the domain information requirements, (2) develop a knowledge resource for unstructured text processing (e.g., for information retrieval or extraction purposes), and (3) eventually use the information in the respective knowledge base for, e.g., considering important cargo constraints when reasoning about proposing a set of possible shipment routes.

Designing and developing an ontology from scratch can be a laborious and time-consuming process. For this reason, there are numerous approaches in learning an ontology in an automatic or semiautomatic way, such as using automatic term extraction and clustering, or information extraction entity and relation extraction [8, 10, 24, 30–33, 45]. In our approach, because of the lack of existing lexical or other knowledge resources in the special cargo domain, we have opted for a top-down method, namely, one that relies on applying knowledge elicitation techniques for acquiring the domain knowledge from the human experts. More details about the size of different components of the ontology are added in Table 1 In this section, we discuss our knowledge elicitation methodology, ontology design, and implementation.

2.1 Methodology and Principles for Ontology Construction

In order to support the planning phase within the special handling cargo sector, a knowledge structure is constructed. Based on an analysis of the ontology life cycle, (dis)advantages, and the conformity to the nature of the special cargo domain, different methodologies are assessed. The result of the analysis of different methodologies and techniques is the augmented UPON (Unified Process for Ontology) methodology [9] with knowledge elicitation and evaluation tools.

The building process of special cargo ontology follows the UPON methodology that is based on a software development process. UPON is augmented with knowledge elicitation techniques to derive knowledge from experts and evaluation techniques to validate the ontology. (Un)Structured interviews including the teach-back method, laddering,[1] and document analysis techniques are implemented

[1] This consists of techniques consists of creating a hierarchy of the gathered knowledge, reviewing, modifying, and validating it together with an expert.

Table 1 Different components of the special cargo ontology

Component	Size
Axiom	724
Logical Axiom	344
Declaration Axiom	197
Declaration Axiom	197
Class	129
Object Property	43
Data Property	20
Individual	7
Annotation Property	4
Class Axiom: SubClassOf	240
DisjointClasses	14
Object Property Axioms: SubObjectPropertyOf	2
InverserObjectProperties	5
FunctionalObjectProperty	8
TransitiveObjectProperty	4
ObjectPropertyDomain	4
ObjectProperyRange	3
Data Property Axioms: FunctionalDataProperty	4
DataPropertyDomain	25
DataPropertyRange	19
Individual Axioms: ClassAssertion	16
Annotation Axiom: AnnotationAssertion	183

into this methodology. The UPON methodology consists of five main workflows, namely, requirements, analysis, design, implementation, and test. In the requirements workflow, the goal is to identify the requirements and desires of the ontology users, which consists of (1) determining the domain of interest and the scope, and (2) defining the purpose, which results in the usage of knowledge elicitation techniques and an Ontology Requirement Specification (ORS) document as well as an application lexicon. In the analysis workflow, different existing ontologies are assessed, and a Unified Modeling Language (UML) use-case diagram is constructed, alongside the application lexicon. In the design workflow, the OPAL (Object, Process, Actor Modeling Language) methodology as well as justification for the relevancy of these concepts to the domain is applied to the concepts. A comprehensive explanation of concepts is defined in this step. The implementation workflow consists of implementing the lexicon and its attributes into Protege and offers performance metrics and visualization of ontology structure. The evaluation of an ontology is crucial and can be done in four strategies: gold standard, application based, data driven, and user based [20]. Due to the lack of gold standard, (technical) application, and data, human assessment is the main reference point. The final workflow is testing the ontology, and this is achieved based on the 'assessment' and 'evaluation' methods. In the assessment method, competence questions and principles are assessed. The evaluation approach consisted of a manual annotation

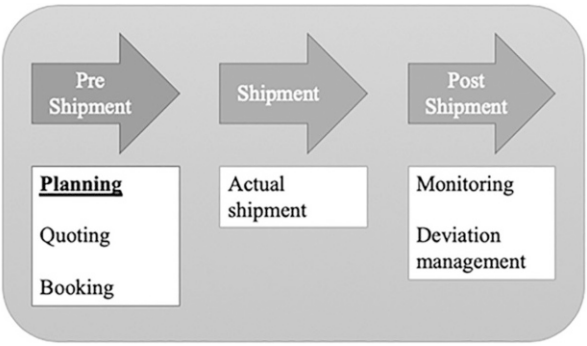

Fig. 2 Components of a cargo shipment

approach to 20 documents that are annotated by an expert. Each phase of this process is explained in more detail in the following sections.

2.2 Requirement Workflow

The application domain of the cargo ontology is the special cargo industry, with a focus on airfreight. This concerns all the processes and products that cover the interactions of special cargo airfreight forwarding within the planning phase of a shipment. Figure 2 shows a general sketch of the components of a (special) cargo shipment. This figure shows the activities that occur before the shipment planning, the actual shipment of the cargo and the activities that occur after the shipment (e.g., management of deviations).

The goal of requirement workflow is to identify the requirements of the ontology users, which consists of '(1) determining the domain of interest and the scope, and (2) defining the purpose' [9]. In this phase, the knowledge engineering techniques are applied according to the CommonKADS method [34] on top of the UPON techniques. The interviews are designed based on the guidelines and samples of CommonKADS. The knowledge elicitation is utilized in three phases, namely, knowledge identification, knowledge specification, and knowledge refinement. Knowledge identification consists of unstructured interviews and document analysis. The next step is the specification of knowledge, with structured interviews. Based on the background knowledge acquired, there are four types of experts: freight forwarders, shippers, GHAs, and support experts. While shippers play a vital role in the transportation of special cargo as it is their products being shipped, they are not concerned with the transportation jargon of the special cargo. Freight forwarders book and arrange the shipments based on the shipper's requirement. Transporters can be separated into the carriers (air carriers) and the handlers (GHA). Due to resource and time constraints, the GHAs are not consulted. The final step of

the requirement workflow is knowledge refinement, and this consists of applying instances and validating the model.

To fulfill the two main goals of this workflow, an Ontology Requirement Specification (ORS) document [36] is derived. The document entails the activities of collecting the special cargo ontology requirements. The cargo that requires special handling is divided into multiple segments, namely, pharma, dangerous goods, perishables, live animals, and high value. In this regard, some information related to the purpose of the special cargo ontology, determination of the available choice set for routing options, including specific product features, capabilities, services of air carriers, and GHAs are found in the ORS document.

Along with the ORS document that includes the competency questions, an application lexicon (based on the knowledge engineering techniques) and a use-case model are the outcomes of this workflow. Applying use-case models based on the competency questions is the final step in the requirement workflow. Figure 3 shows the visualization of this use case. Laddering is conducted with a support expert and is used to elicit the UML diagram.

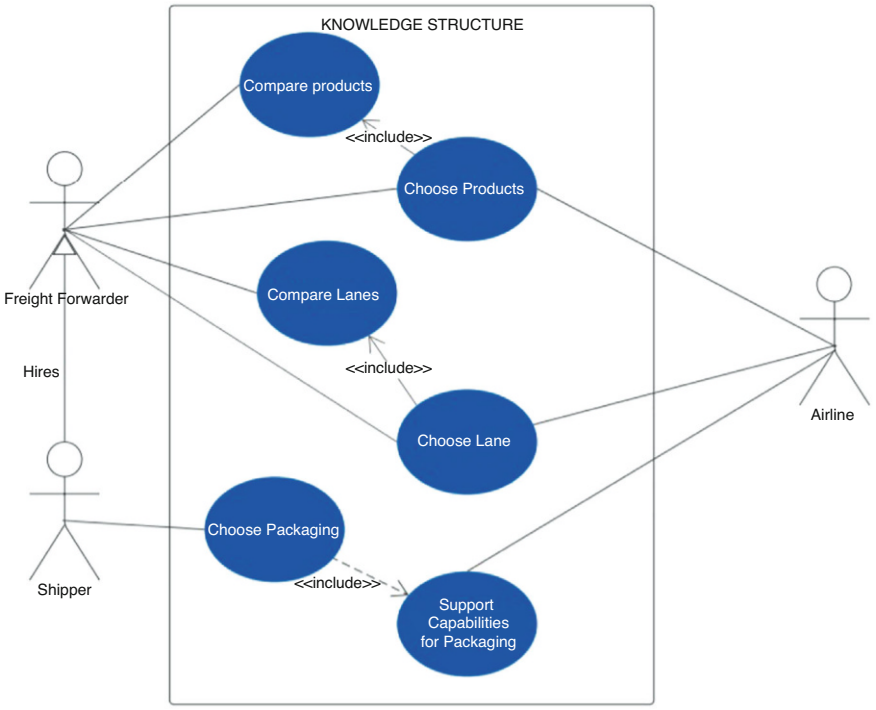

Fig. 3 Cargo ontology use case

2.3 Analysis Workflow

The analysis phase aims to refine and structure the identified requirements of the previous step. This includes reusing existing resources, modeling the application scenario using UML diagrams, and building the glossary. Considering the reuse of existing resources also entails the assessment of other domain ontologies. Existing resources or ontologies have been acquired through a search of several Ontology Libraries (OL). IATA—ONE Record,[2] the NASA Air Traffic Management Ontology,[3] and the Air Travel Booking Ontology[4] are assessed for the relevance to the domain of built cargo ontology.

The IATA—ONE Record ontology, The NASA Air Traffic Management (ATM) Ontology, and The Air Travel Booking Ontology are implemented in a message system, air traffic management, and air travel booking service, respectively. They are used in different stages of the process (i.e., planning vs booking), and the domains are not completely compatible. Although in the context of the Semantic Web, ontologies are often used with a purpose different from the original creators of the ontology [36], and these ontologies do not offer significant benefits to be implemented or associated with the Special Cargo Domain.

The next step is to model the application scenario based on the drafted UML use-case diagram, in the form of a simple UML class diagram. A part of this diagram is shown in Fig. 4, as the result of the elicitation technique laddering. The final step of the analysis workflow is to build the first version of the glossary concerning the concepts of the domain, which will merge the application lexicon and the domain lexicon.

2.4 Design Workflow

The identified entities, actors, and processes and the relations among them in the previous workflow are refined in the design phase. The steps within this workflow consist of inhabiting, categorizing the concepts according to the OPAL methodology [42], and refining the concepts and their relations. OPAL is organized into three primary modeling aspects: actor, processes, and object. The identification of the OPAL methodology, as well as a justification of why such entities exist in the ontology, is defined under the lexicon. The subclasses are related to the main class through a ‘kind-of’ or an ‘is-a’ relation. When a ‘part-of’ relation is defined, it is found in column ‘notes’. The object, data properties, and the related explanation are found in the ontology.

[2] https://www.iata.org/en/programs/cargo/e/one-record.

[3] https://data.nasa.gov/ontologies/atmonto/ATM.

[4] https://www.southampton.ac.uk/~cd8e10/airtravelbookingontology.owl.

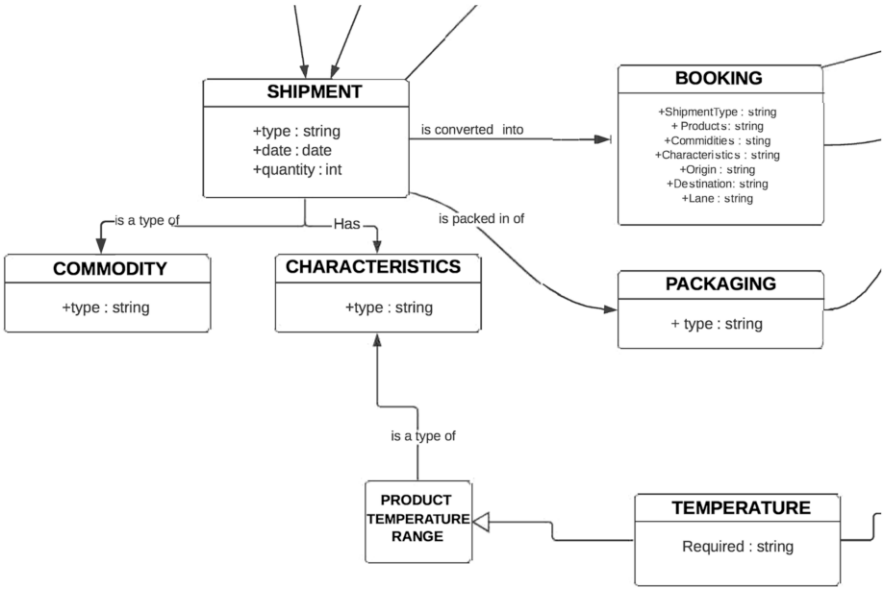

Fig. 4 A part of the Cargo Ontology class

2.5 Implementation Workflow

In this phase, ontology is formalized in a language and implemented with regard to its components. The special cargo ontology is constructed in Protege and written in RDF (Resource Description Framework) and OWL (Ontology Web Language). A part of the visualization of the special cargo ontology is shown in Fig. 5.

2.6 Test Workflow

While each ontology differs in structure and domain, testing is vital to assess the domain compliance. The goal of the test phase is to evaluate the ontology and its components and requirements. The evaluation is performed based on human-based and task-based assessment. Human-based assessment is divided into two parts: the competency questions and the principles assessment. The competency questions (CQ) are drafted in the requirement workflow, as the manual assessment will be based on the CQChecker module of Bezerra et al. [27]. The principle assessment is a subjective tool, which requires the collaboration of the ontology engineer and a

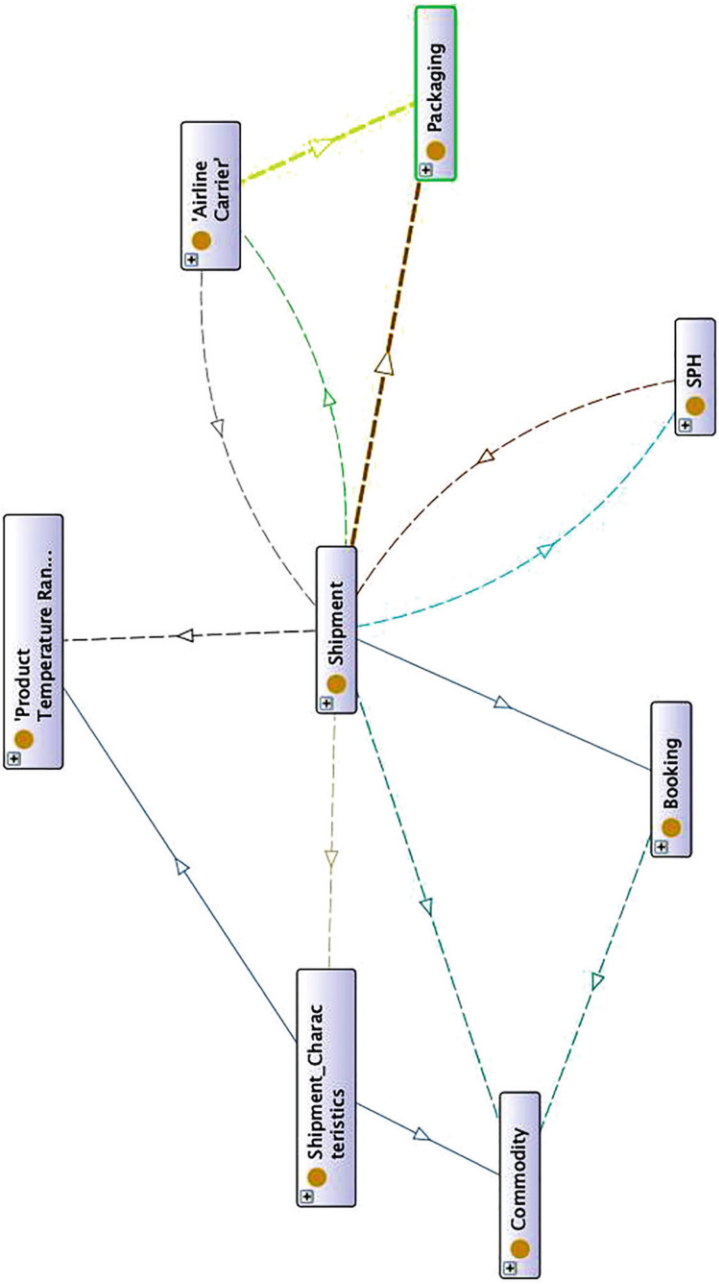

Fig. 5 'Shipment' concept in the Cargo Ontology class

Table 2 Competence Questions

Question	Real-life answer	Ontology answer	Compliance and relation
Does a pharma solution have a booked temperature range?	Yes	Yes	YES: Temperature controlled solutions 'has temperature range' some booked temperature range
Does lithium batteries transport have restrictions?	Yes	Not fully deductible	SEMI: Dangerous goods class 'has maximum capacity' (classes are not populated yet)

Table 3 Design principles

General design principles compliance	Compliance
The design should clearly state its purpose, so the user knows what the design has to offer to avoid unclear expectations	Compliant. During the extent of this research, the scope, the domain, and its purpose have been defined as well as the expectations by the LARA project
The design should remain its stability throughout time, changes, and additions	Compliant, so far. As the ontology is constructed as of late, time is hard to test on this design. However, similarly to the maintainable design principle, Protégé allows for adjustment and augmentation

domain expert. Tables 2 and 3 show some parts of these two different assessments, competence questions and design principles, respectively.

2.7 Evaluation Workflow

Task-based, data-driven evaluation is conducted by a domain expert. The evaluation is executed on two sets of ten documents concerning special cargo, collected from online cargo websites and news articles. The expert who annotated the documents has experience within the freight forwarding process as well as the risk analysis of lanes.

The final step in the testing phase is to adjust the ontology according to the result of the overall evaluation. There were three concepts ('Certification', 'Hub', 'Documentation') that were neglected in the original ontology which were implemented after the evaluation. In Table 4, a snippet from this evaluation is shown. During the analysis of the evaluation, it became clear that certain small or significant attributes were omitted in the process of creating the ontology, or in return some attributes were insignificant. In the result of this analysis, these attributes were omitted or inserted.

Table 4 A snippet from cargo ontology evaluation

Annotation	Relevance	Presence
Our products allow you to get your life-saving cargo to its destination	Yes	Yes, triple incorporated: product, pharma, and the relation
Cool Center	Yes	Yes, cool center is a synonym for temperature-controlled environment, this concept is incorporated
Highly trained experts can stand by 24/7/365 to monitor and support	Yes	Yes, trained personnel and monitoring are incorporated

2.8 Summary

In the LARA project, knowledge representation is developed for the special handling goods and services in the airfreight sector. It is designed based on the software engineering methodology with the aim of digitizing the determination of the choice set of solutions and routes for the airfreight forwarders by making data transparent and understandable to machines. For the integration of disparate knowledge sources, a special cargo domain ontology of shipping concepts is constructed for the domain of goods transported by air in a semiautomatic manner. As a structured resource, the special cargo ontology provides valuable insights into the scope of the application, the different components of the system, and the interaction between them. It can be used during the actual operation of the system [6, 22]. As an example, the fact that consumer-ready laptop computers contain a lithium battery can be modeled in the ontology means that when processing a request for shipping laptops, the system can determine that the cargo service needs to allow for lithium batteries to be shipped.

The UPON methodology is used for the construction of the cargo domain knowledge structure to get the relevant concepts and attributes. This output is evaluated based on reviewed evaluation methods and adjusted accordingly. The ontology is integrated into a software program to obtain an applicable product of the special cargo scope and domain, and subsequently, the final product is the base of an artificial intelligence route advisor based on the semantic web for the special cargo sector.

3 Case Study: Lane Analysis and Route Advisor

In the past few decades, international freight transportation has increased rapidly. This rise can be explained by technological developments, simplifying the global transport process and causing a decline in shipping costs [37]. It has led to a growing demand for freight forwarding services. Freight forwarding companies can be hired to handle the logistics of shipping goods from the customer to the consignee.

However, the process of transportation carries many risks, for which the freight forwarding company has to take responsibility. Certain types of cargo may require

strict conditions during transit. For instance, some pharmaceutical products are temperature-sensitive and have to be kept at a specific temperature throughout the entire process. When constructing the route the cargo should follow, the type of packaging and possible exposure to external weather conditions need to be taken into account. Furthermore, when transporting high-value cargo (HVC) like electronics, the number of crime incidents increases. Hence, additional security measurements should be taken into consideration for HVC goods.

Freight forwarding companies aim to maintain high customer satisfaction as satisfied customers will presumably hire the company again and might help recruit other customers through positive feedback[28]. Key elements driving customer satisfaction are the service quality and the perceived value [15]. Hence, to avoid incidents and thus increase customer satisfaction, it is essential to develop a risk assessment model and to determine high-risk lanes.

Despite the aim of freight forwarders to work as carefully and efficiently as possible, incidents are inevitable. While considerable amounts of data are available regarding every incident, a lot of potential still exists to gain knowledge on factors that cause (or contribute to) incidents. Research on this matter is essential for freight forwarders; the prevention of incidents can not only contribute to keep costs as low as possible but also help forwarders maintain their reputation of a reliable forwarder.

The question arises whether factors or even combinations of factors exist that drive incident risk. A comprehensive study concerning the incident data is needed to answer this question, which is the objective of this research. This chapter focuses on incident analysis and tries to determine which factors drive risk.

For this research, high-dimensional data on incidents was provided by one of the major freight forwarding companies in the industry.

4 Natural Language Processing for Incident Handling

Logistics is defined as the process of planning, implementing, and controlling procedures for the efficient and effective transportation and storage of goods including services, and related information from the point of origin to the point of consumption for the purpose of conforming to customer requirements [25].

With regards to logistics, the data focuses on the transportation of cargo, specifically incident handling with regard to air cargo. Because this is the case, it is interesting to look at ways other chapters tackled this issue of cargo risk assessment. One of these risks is cargo loss. This is defined by Wu et al. [44] as either cargo damage or cargo theft.

According to [29], cargo damage is the most occurring problem in the logistics sector. These authors mention five main causes of cargo damage: human error (such as miscommunication); handling error (examples include incorrect placement in plane or having incorrect/missing documents); machine/tool error (such as having old or broken equipment); environment (such as temperature); and packing material.

When it comes to cargo theft, [26] mentions that employees, as well as outside offenders, may steal cargo. These authors also mention a couple of reasons why it is often difficult to detect cargo theft. One of these reasons is the fact that thefts are often under-reported. They further mention ways in which the amount of cargo theft can be decreased. These methods include, but are not limited to, placing containers with doors facing each other (so that it is more difficult to remove cargo) and minimizing waiting times for vehicles (because it is easier to steal from a still-standing vehicle).

These issues have been tackled by other authors using both predictive and descriptive analysis techniques to gain insight on cargo loss [44]. One thing they found is that high-value cargo should not be sent as land cargo, and to certain regions not as sea cargo either.

While the above-mentioned results are interesting, Hazen [18] mentions a few reasons why data analysis results regarding the logistics and supply chain industries should be considered carefully. Most of these reasons are due to data quality issues. According to these authors, data in this sector is often full of errors. They mention four key attributes in data quality that could use improvement: accuracy, timeliness, consistency, and completeness.

4.1 Random Forest Decision Trees

One of the issues that the company has been dealing with is data quality. Registration of an incident may provide text describing the incident. The classification of incidents is a subjective process in which mistakes can happen. Comparing text could, therefore, be a competent way to eliminate these mistakes. A suitable method is to process the data using Natural Language Processing (NLP) and to classify the incident using a classification model with this processed data.

NLP is a subfield of computer science that uses computational techniques to learn, understand and produce human language content [19]. The authors mention multiple reasons why NLP is useful. Among them are translating and helping the human-machine communication, which are both relevant for this research. Also, the process of analysis and learning from human language content which is available online is discussed in this chapter and might be relevant for this research as well.

One way to analyze and learn from human language content is by using machine learning algorithms. Random forests can be seen as a combination of multiple tree predictors in which each tree depends on the values of a random vector sampled independently and each following the same distribution for all trees that are included in the forest [7]. Although different types of trees exist, in this case, decision trees are used. In decision trees, the decisions are the edges of the tree and form the nodes for data classification. Decision trees are applied commonly in machine learning; one reason for their frequent usage is that they are easy to interpret [43].

However, a random forest algorithm is generally preferable over using just a decision tree; random forests improve performance by training multiple decision trees [43]. These trees are chosen randomly because, in that way, the chance of correlation between individual trees is reduced, and more accurate results are obtained.

4.2 Implementation

There are several NLP environments on the market. One of the more standard environments is the Natural Language Toolkit (`NLTK`) which is combined with the python `sklearn` library for the best results. A common process in NLP is tokenization. In this process, sentences are broken up into individual words, where any capital letters and interpunction are also removed. The classification model then uses a set number of most common words in the incidents. Using these most common words, the random forest classification is then trained on the training data using a set number of classification trees.

Random forest classification uses a trained classification model that is capable of classifying data based on processed text. It requires the incident data to be split up into a training and a testing set. Because of the large number of incidents, it is possible to split them up into a training set that contains 75% of the incidents and a testing set that includes 25% of the incidents. To make the random forest classification easier, the problem is reduced to a single classification problem. This means that all possible combinations of levels are given an ID, which the model tries to classify.

The accuracy of the classification model will vary based on the chosen parameters during the NLP and the classification process. Furthermore, it would be extra interesting to look at the incidents that were wrongly classified, as the original classification by the incident handlers could also be wrong.

4.3 Results and Discussion

Natural Language Toolkit (`NLTK`) combined with a random forest classifier provides the prediction accuracy as shown in Table 5. The NLP random forest algorithm was implemented with 1, 10, and 100 trees.

Table 5 NLP results based on NLTK with random forest classifier

#trees	Precision
1	82.0
10	82.6
100	83

An interesting result is that in all cases, the precision rounds up to being an integer, and after 100 trees, the precision does not rise much more than that integer. Since adding random forests are not prone to over-fitting and the precision flattens out at 100 trees, 100 trees seem like an adequate number of trees.

It could be interesting to look at incident types that are predicted wrong relatively often and see if it would not be better to put these under a different class.

5 Statistics and Machine Learning to Improve Risk Assessment

To be able to determine possible factors that drive incident risk, a multinomial classification model was implemented on the incident data. In this model, features are defined for every incident type that significantly predict this type. After literature research, it became clear that a Logistic Regression Model suited the data well.

5.1 Logistic Regression

A classification method known as the Logistic Regression model is used during the analysis. Regression models are used to calculate the interdependency between an outcome (response variable) and the variables thought to affect this outcome (explanatory variables). The most simple form of a regression model is the linear regression model, in which a linear function is mapped between data points. The logistic regression model is in certain ways similar to linear regression, but there are a few differences. The main difference is that with the logistic regression (logit) model used in this research, the outcome variable is discrete instead of continuous. The statistical model is typically estimated via (simulated) maximum likelihood estimation [21, 38]. Logit family models are widely applied in the transportation domain [2, 14]. Examples include: mode choice [11], route choice [23], choice of departure time [41], location choice [3, 40], and choice of products and services [13, 39].

In the logistic regression model, the relationship between the response variable and explanatory variables is expressed as a simple equation:

$$g\left(\mathrm{E}[Y]\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \tag{1}$$

where g is the logit link function. In the equation above, the αrepresents a constant term, the $x_i$represent the explanatory variables, and the $\beta_i$represent a measure of the degree to which the response variable is explained by variable x_i[17]. For every explanatory variable, a t-test is performed on the corresponding $\beta$to test whether it can be statistically proven that the explanatory variable influences the response

variable. To determine how useful the explanatory variables are in predicting the response variables, the ρ^2 statistic or a likelihood ratio test can be used [4].

For the implementation of the Logistic Regression in a Machine Learning fashion. a procedure is required that identifies features that are of importance to the response variable. A procedure that determines this is called Recursive Feature Elimination (RFE). After training the classifier and computing the ranking, the feature with the smallest ranking criterion is removed [16]. This step is repeated until a certain number of features n remains.

5.2 Methodology

To classify incidents into categories, these categories first had to be determined. The data that was provided for this analysis contained a feature that described what kind of incident happened. Based on this column, the incident types with the highest number of occurrences were chosen to implement in the model. Also, types that were prioritized by the company, but did not have a significantly high number of occurrences, were taken into account. In total, nine categories were obtained.

Two classification models were used, using different python packages: `Biogeme` and `RFE`. Both use the incident types described above as categories. The methods per model are described in the following subsections.

5.3 Statistical Implementation

For the implementation of the statistical Logistic Regression model, a package called `Biogeme` by Bierlaire [5] was used. The model performs a multinomial classification and determines significant features that predict all possible classes (the incident types). For this model, it was necessary to manually determine which variables drive the chosen incident types most and include these in the model. To determine these variables cross-tabular matrices were used, which show the proportional relation of a variable to a particular incident type. More specifically, they depict which possible values of certain variables show a connection to an incident type. The cross tab had all incident types as rows and all possible values for the variable to take into consideration as columns. For example, when taking a region variable into account, a certain incident type might have a strong correlation with a specific region. In this case, a binary variable was created, where '1' equals the situation where the incident was reported in that region, and 0 otherwise. Next, this variable was added to the regression model for this incident type. In this process, also the number of occurrences per region has to be taken into account, to avoid an unreliable view on possible predictors. If a certain region only occurred once in the data, and by coincidence an incident occurred on the shipment connected to this region, the percentage error will be 100%. Therefore, a threshold was set for the

cross tabs, where every possible value taken into account in the cross tab had to have at least a minimum number of occurrences in the incident data.

The chosen features were entered for the corresponding incident type. The python package `Biogeme` was able to check for all features whether it was an important predictor for the incident type. After the first run, a base model was created that included all features for which the value of the t-test statistic was bigger than 1. This does not imply all these values are statistically significant on a 5% level, but they have enough descriptive purpose for the model. After this, all features included in the base model were analyzed for possible combinations of features with a high correlation. For instance, when the incident was caused in a particular city, the corresponding country at fault is of course always the country containing that city. This collinearity has a negative effect on the performance of the model, so for the combinations of features with high correlation, the less specific features are excluded. In the example, it would be more important to look at the city specifically, than to only take the country into consideration. Hence, in such case, the country is excluded from the model features.

Using this base, all other features were checked again for possible relevance to the incident type. This was done by performing a batch run on the base model, where every time one of the features not contained in the base model was added to the base. The rho squared and likelihood ratio test results were compared for all resulting models to determine the optimal one.

In a general logistic regression model, the predictive power of a feature cannot be determined by the beta value, since the influence on the utility function is defined as the beta times the value of the feature. Therefore, a feature with high values generally has a smaller beta than a feature with low values. However, since the features used in the regression model are all dummy variables (so only binary values are included), it is possible to compare the strength of the feature on the beta value.

5.4 Recursive Feature Elimination

For comparison with the statistical Logistic Regression model, a machine learning Logistic Regression was used as an alternative. For implementation, one of the more general packages and approaches within python, known as `sklearn`, provided the necessary tools. The main difference with the statistical approach concerns the selection of possible values per feature. All features that were determined of importance were used for the machine learning approach as well, but instead of selecting important values per feature manually using cross tabs, these values were decided using Recursive Feature Elimination (`RFE`). A data set was created in a 'one-hot-encoding' fashion where all features were split up into binary variables. So, for example, all possible values for the party responsible for the incident in the data got their own column, where 1 indicates that the incident was caused by that specific company and 0 indicates that another company was responsible. Since some columns have many different possible values, and some of these values only occur a

few times in the data, it was decided to set a threshold on the number of occurrences. This was done to ensure predictive power and reduce the running time of the model.

It was impossible to run the model as a multinomial regression model, since then the overall best features were decided for all incident types in total, instead of per type. Therefore, the formulation of the model changed to a set of binary decisions: the model was run for every incident type separately. To achieve this, a binary column per incident type was added to use as classification feature. A drawback of this implementation is that it can also produce negative betas. Because the model is run for every incident type separately, it is now classifying features on the constraint 'is it a predictor for incident type X or not', instead of taking into account all incident types. A negative beta shows that the corresponding feature is not a good predictor for incident type X, but it is a significant predictor for some other incident type.

The machine learning implementation works in the following way. First, the model splits the data set into a training and testing set of 70% versus 30%. `RFE` calculates the best features to be used by the model using the training set. It was decided to let the `RFE` determine ten features per incident type. Following the `RFE`, a simple logistic regression model is constructed, using the training set to fit the model, which then provides results based on the testing set. The accuracy of the model shows the percentage of the prediction by the model that was correct.

5.5 Results

The analysis of features that could possibly predict certain incident types has led to a little over 300 different variables among the nine different regression models. So, on average around 30 possible predictors were determined per incident type. The Logistic Regression models determined the features that were the most important predictors per incident type.

The classification model implemented with the python package `Biogeme` gives as output an overview of all features used. A statistical test is conducted for every β, which shows whether the influence of the corresponding explanatory variable is statistically significant for the incident type it was tested for. As explained in the chapter 'Methods', a base model was created with all features for which the t-test value was bigger than 1. Per feature and value combination, the Beta gives a measure of how much this combination influences the incident type. Thus, the features with the highest Beta values for each incident type are the strongest predictors. The p-value depicts the significance of the feature in the base model. The smaller the p-value, the higher the accuracy of this feature for the model. After running the base model separately from the full model, some insignificant p-values were obtained, while they were significant when all of the features were taken into account. It would have been better to iterate over the results of the model and to exclude the insignificant features every time. However, due to the immense running time of the model, it was decided to focus on this base model and to accept the few high p-

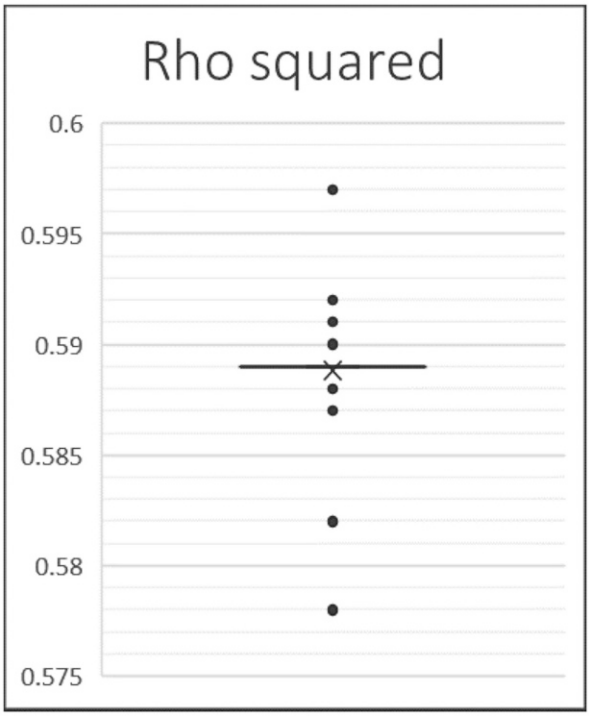

Fig. 6 All rhos of batch run

values. The rho squared for the base model was equal to 0.575 and the likelihood ratio test was equal to 93,295.63.

The batch run of the base with every single feature separately produced 220 models. Figure 6 shows a box plot of all rho squares that were obtained per model. Figure 7 shows all likelihood ratio test values. The highest rho squared and likelihood ratio are equal to 0.597 and 97,203.09, respectively.

The Logistic Regression model performed with the columns detected by the `RFE` gave as output ten features with most predictive power per incident type. The accuracy of each model can be found in Table 6.

5.6 Discussion

As mentioned before, the results show some insignificant p-values in the base model. These values could be explained by the fact that the corresponding features occur often for multiple incident types. Thus, they have a substantial variance as a predictor. Therefore, these features should be considered with caution. It should also be noted that some of the incident types that were implemented in the model

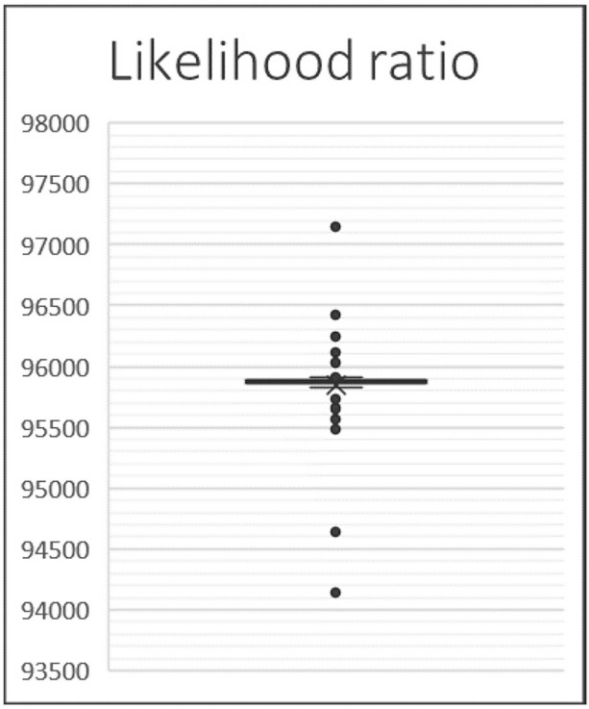

Fig. 7 All likelihood ratios of batch run

Table 6 Accuracy for the Logistic Regression model based on RFE determined features

Incident type	Accuracy
A	0.904
B	0.713
C	0.987
D	0.867
E	0.951
F	0.978
G	0.965
H	0.987
I	0.999

did not have many occurrences in the data. After running the full model, only a few predictors were determined for the base model, and they were all deemed insignificant after running the base model on its own. Therefore, a high number of occurrences is needed per incident type to acquire significant results.

The box plots with the results of the batch run (Figs. 6 and 7) show that no significant differences could be found between the resulting models. The rho squared has its mean at 0.589, and only has a few outliers. Still, the base model had a rho squared of 0.575, so adding another variable to the model generally leads to

a better performing model. The highest rho squared and likelihood ratio were 0.597 and 97,203.09, respectively. However, adding the binary variable that achieved these maximum statistical measures lead to collinearity with other features. The features corresponding to the next few highest rho squares lead to the same issue. After the sixth feature, the rho squared becomes 0.59 for any feature added to the model (except for a few). Therefore, the batch run does not provide any significant results of features to add.

5.7 Comparison of the Statistical and RFE Models

The results of the two models implemented by `Biogeme` and `RFE` require special attention to compare. A reason for this is the fact that the features of the `Biogeme` model are judged by the statistical t-test that is performed and the resulting p-values. However, the machine learning-based model does not judge the performance of features on a statistical test. Instead, it splits the data up into a train and a test set, trains the model on the train set, and calculates the performance of the model based on the test set.

Another challenge with the comparison is the negative betas that occur for the `RFE` implementation. The statistical model is a multinomial implementation, which implies that it runs the model on all nine different incident types at once, and determines the appropriate features accordingly. The formulation of the machine learning model, however, is a separate set of binomial decisions. For every incident type, the model is run separately, in order to find features per incident type. This means that the model is classifying features on the constraint 'is it a predictor for incident type X or not', instead of taking into account all incident types. Because of this, the results for the `RFE` model contain negative betas. A negative beta shows that the corresponding feature is not a good predictor for incident type X, but it is a significant predictor for some other incident type. However, for the analysis conducted in this report, the negative betas do not add any important information. Hence, only the positive betas should be taken into account.

Still, when comparing the significant features per incident type, most of the features with a positive beta in the machine learning implementation also occurred in the results for the statistical implementation. This shows that these values are in fact important predictors for the incident types.

6 Summary, Challenges, and Conclusion

This research project is directly related to BDVA SRIA's strategic and specific goals, particularly to the topic Data Analytics to improve data understanding and providing optimized architectures for analytics of data-at-rest and data-in-motion. In this project, we conduct research into solutions based on advanced data analytics

that combine the integration of various data sources ('big data'), AI-based methods such as machine learning, and natural language processing for prescriptive analytics and decision making. These methods can be applied to the optimization of route planning in global transportation and freight forwarding of sensitive products with special handling needs (e.g., COVID-19 vaccine) targeted at air freight shipment.

According to the European Big Data Value Strategic Research and Innovation Agenda (SRIA) [32], understanding data has been one of the greatest challenges for data analytics. In this regard, we use semantic and knowledge-based analysis specifically ontology engineering for Big Data sources in the special cargo domain to improve the analysis of data and provide a near-real-time interpretation of the data (i.e., accurate prediction of the lane performance). Moreover, employing Big Data analytics we develop an ontology for the products and services offered for air freight logistics providers. Based on this, a search engine can be developed to determine the available routing options for a shipment with specific features. Thus, it provides additional value in the transportation sector, leads to more efficient and accurate processes, and improves operational efficiencies and customer service.

The work has some limitations and challenges. Evaluation of the special cargo ontology is difficult and needs manual intervention, which is time-consuming and subjective. Expert intervention is required at every step of constructing ontology. Nevertheless, this work aims to make a significant contribution to the digitization of global freight forwarding, which may also pave the way toward 'no-touch' planning in airfreight transportation.

In this chapter, we also present a case study applying a novel palate of data analytics for risk assessment. A natural language processing classification model used on text in the incident handling data shows at least an 82% accuracy at identifying incident types. Furthermore, via a statistical logistic regression model for classification, it can be proven that several features are significant predictors of certain incident types. A machine learning logistic regression model also identified similar features. Focusing on these features can help the company prevent similar incidents in air cargo handling in the future.

The chapter addresses some important challenges of the airfreight industry for shipping goods with special handling needs such as vaccines. In order to design, develop, and optimize the decision making of a routing service in the special cargo domain, it is necessary to conceptualize and structure the available knowledge from different resources as a special cargo knowledge resource (ontology). This ontology is efficient for reasoning and can be used during the actual operation of the system. As an example, the fact that vaccines must be stored at an ultracold temperature can be modeled in the ontology, which means that, when processing a request for shipping vaccines, the system can determine that the cargo service needs to allow the shipment of products with special temperature needs.

Using the special cargo ontology, more heterogeneous sources of information can be automatically extracted and integrated. This information includes, for example, previous incidents and service performance. This knowledge base can be used in various downstream tasks, e.g., risk assessment model as a feature-extraction source. On the other hand, the machine learning algorithms applied in

risk assessment tasks can be used for enriching the cargo domain ontology and map the extracted information to the structured knowledge source.

This research is directly related to TKI Dinalog's innovation roadmap, specifically to the topic Advanced Data Analytics in Transport Planning within the Smart ICT Roadmap.

Acknowledgments The work is supported by TKI Dinalog, the Dutch Institute for Advanced Logistics, for the LARA project—Lane Analysis & Route Advisor (Project reference number: 2018-2-171TKI). TKI Dinalog is the Knowledge and Innovation Partnership in which business, knowledge institutes, and government work together in the innovation program of the Dutch Topsector Logistics. We would also gratefully like to acknowledge the support of Daniel Lutz and Carolin Heinig in the case study application on incident handling and risk assessment.

References

1. Asim, M., Wasim, M., Khan, M. U., Mahmood, W., & Abbasi, H. M. (2018). A survey of ontology learning techniques and applications. *Database: The Journal of Biological Databases and Curation*. https://pubmed.ncbi.nlm.nih.gov/30295720/.
2. Ben-Akiva, M., & Bierlaire, M. (2003). Discrete choice models with applications to departure time and route choice. In *Handbook of Transportation Science* (pp. 7–37). Kluwer.
3. Berkelmans, G., Berkelmans, W., Piersma, N., van der Mei, R., & Dugundji, E. R. (2018). Predicting electric vehicle charging demand using mixed generalized extreme value models with panel effects. *Procedia Computer Science*, *130*, 549–556.
4. Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. *Critical Care*, *9*, 112–118.
5. Bierlaire, M. (2016). *Pythonbiogeme: a short introduction*. Tech. rep., EPFL, Switzerland.
6. Black, W. J., Jowett, S., Mavroudakis, T., McNaught, J., Theodoulidis, B., Vasilakopoulos, A., Zarri, G. P., & Zervanou, K. (2004). Ontology-enablement of a system for semantic annotation of digital documents. In *SemAnnot@ ISWC*.
7. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
8. Buitelaar, P., Cimiano, P., & Magnini, B. (2005). *Ontology learning from text: methods, evaluation and applications* (Vol. 123). IOS Press.
9. De Nicola, A., Missikoff, M., & Navigli, R. (2005). A proposal for a unified process for ontology building. In *International Conference on Database and Expert Systems Applications* (pp. 655–664). Springer.
10. Drymonas, E., Zervanou, K., & Petrakis, E. G. M. (2010). Unsupervised ontology acquisition from plain texts: The ontogain system. In C. J. Hopfe, Y. Rezgui, E. Métais, A. Preece, & H. Li (Eds.), *Natural Language Processing and Information Systems* (pp. 277–287). Springer.
11. Dugundji, E. R., & Walker, J. L. (2005). Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. *Transportation Research Record*, *1921*(1), 70–78.
12. Faghih-Roohi, S., Akcay, A., Zhang, Y., Shekarian, E., & de Jong, E. (2020). A group risk assessment approach for the selection of pharmaceutical product shipping lanes. *International Journal of Production Economics*, *229*, 107774.
13. Feilzer, J. W., Stroosnier, D., Koch, T., & Dugundji, E. R. (2021). Predicting lessee switch behavior using logit models. *Procedia Computer Science*, *184*, 380–387.
14. Garrow, L. (2016). *Discrete choice modelling and air travel demand*. Routledge.

15. Gil-Saura, I., Berenguer-Contri, G., & Ruiz-Molina, E. (2018). Satisfaction and loyalty in b2b relationships in the freight forwarding industry: adding perceived value and service quality into equation. *Transport*, *33*(5), 1184–1195.
16. Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*, 389–422.
17. Hall, G., & Round, A. (1994). Logistic regression – explanation and use. *Journal of the Royal College of Physicians of London*, *28*(3), 242–246.
18. Hazen, B. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72–80.
19. Hirschberch, J., & Manning, C. (2015). Advances in natural language processing. *Science Magazine*, *349*, 1184–1195.
20. Hlomani, H., & Stacey, D. (2014). Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web Journal*, *1*(5), 1–11.
21. Hosmer, D., & Lemeshow, S. (2013). *Applied logistic regression*. John Wiley and Sons.
22. Klein, W., Zervanou, K., Koolen, M., van den Hooff, P., Wiering, F., Alink, W., & Pieters, T. (2017). Creating time capsules for historical research in the early modern period: Reconstructing trajectories of plant medicines. In M. Hasanuzzaman, A. Jatowt, G. Dias, M. Düring, & A. van den Bosch (Eds.), *Proceedings of the 4th International Workshop on Computational History (HistoInformatics 2017) co-located with the 26th ACM International Conference on Information and Knowledge Management (CIKM 2017), Singapore, November 6, 2017*, *CEUR Workshop Proceedings* (Vol. 1992, pp. 2–9). CEUR-WS.org. http://ceur-ws.org/Vol-1992/paper_2.pdf
23. Koch, T., & Dugundi, E. R. (2021). Limitations of recursive logit for inverse reinforcement learning of bicycle route choice behavior in Amsterdam. *Procedia Computer Science*, *184*, 492–499.
24. Lubani, M., Noah, S. A. M., & Mahmud, R. (2019). Ontology population: approaches and design aspects. *Journal of Information Science*, *45*(4), 502–515.
25. Mangan, J., & Lalwani, C. (2016). *Global logistics and supply chain management*, 3rd ed. Wiley.
26. Mayhem, C. (2001). The detection and prevention of cargo theft. *Trends and Issues in Crime and Criminal Justice*, *214*, 1–6.
27. Missikoff, M., & Taglino, F. (2002). Business and enterprise ontology management with symontox. In *International Semantic Web Conference* (pp. 442–447). Springer.
28. Naumann, E., Williams, P., & Khan, M. (2009). Customer satisfaction and loyalty in b2b services: directions for future research. *The Marketing Review*, *9*(4), 319—333.
29. Oktaviani, N., Yadia, Z., Nasution, N., & Veronica, V. (2017). How to reduce cargo damage. *Advances in Engineering Research*, *147*, 661–670.
30. Reshadat, V., & Faili, H. (2019). A new open information extraction system using sentence difficulty estimation. *Computing and Informatics*, *38*(4), 986–1008.
31. Reshadat, V., & Feizi-Derakhshi, M. R. (2012). Studying of semantic similarity methods in ontology. *Research Journal of Applied Sciences, Engineering and Technology*, *4*(12), 1815–1821.
32. Reshadat, V., Hoorali, M., & Faili, H. (2016). A hybrid method for open information extraction based on shallow and deep linguistic analysis. *Interdisciplinary Information Sciences*, *22*(1), 87–100.
33. Reshadat, V., HoorAli, M. & Faili, H., (2019). A new method for improving computational cost of open information extraction systems using log-linear model. *Signal and Data Processing*, *16*(1), 3–20.
34. Schreiber, A. T., Schreiber, G., Akkermans, H., Anjewierden, A., Shadbolt, N., de Hoog, R., Van de Velde, W., Nigel, R., Wielinga, B., et al. (2000). *Knowledge engineering and management: the CommonKADS methodology*. MIT Press.

35. Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, *25*(1–2), 161–197. http://dx.doi.org/10.1016/S0169-023X(97)00056-6
36. Suárez-Figueroa, M. C., Gómez-Pérez, A., Villazón-Terrazas, B. (2009). How to write and use the ontology requirements specification document. In *OTM Confederated International Conferences on the Move to Meaningful Internet Systems* (pp. 966–982). Springer.
37. Tester, K. (2017). The impact of technological change on the shipping industry. Technology in Shipping (pp. 11–20).
38. Train, K. (2003). *Discrete choice methods with simulation*. Cambridge University Press.
39. van Kampen, J., Pauwels, E., van der Mei, R., & Dugundji, E. R. (2019). Analyzing potential age cohort effects in car ownership and residential location in the metropolitan region of Amsterdam. *Procedia Computer Science*, *151*, 543–550.
40. van Kampen, J., Pauwels, E., van der Mei, R., & Dugundji, E. R. (2021). Understanding the relation between travel duration and station choice behavior of cyclists in the metropolitan region of Amsterdam. *Journal of Ambient Intelligence and Humanized Computing*, *12*(1), 137–145.
41. Vegelien, A. G., & Dugundji, E. R. (2018). A revealed preference time of day model for departure time of delivery trucks in The Netherlands. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1770–1774). IEEE.
42. Vrandečić, D. (2009). Ontology evaluation. In *Handbook on ontologies* (pp. 293–313). Springer.
43. Wu, J., Feng, T., Naehrig, M., & Lauter, K. (2016). Privately evaluating decision trees and random forests. *Proceedings on Privacy Enhancing Technologies*, *2016*(4), 335–355.
44. Wu, P., Chen, M., & Tsau, C. (2017). The data-driven analytics for investigating cargo loss in logistics systems. *International Journal of Physical Distribution & Logistics*, *47*(1), 68–84.
45. Zervanou, K., Korkontzelos, I., van den Bosch, A., & Ananiadou, S. (2011). Enrichment and structuring of archival description metadata. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (pp. 44–53). Association for Computational Linguistics, Portland, OR, USA. https://www.aclweb.org/anthology/W11-1507
46. Zillner, S., Bisset, D., Milano, M., Curry, E., García Robles, A., Hahn, T., Irgens, M., Lafrenz, R., Liepert, B., O'Sullivan, B., & Smeulders, A. (Eds.), *Strategic research, innovation and deployment agenda: AI, data and robotics partnership* (3rd ed.) BDVA, euRobotics, ELLIS, EurAI and CLAIRE (2020). https://ai-data-robotics-partnership.eu/wp-content/uploads/2020/09/AI-Data-Robotics-Partnership-SRIDA-V3.0.pdf
47. Zillner, S., Curry, E., Metzger, A., Auer, S., & Seidl, R. (2017). *European big data value strategic research & innovation agenda*. Big Data Value Association.