Supply Chain Visibility with Linked Open Data for Supply Chain Risk Analysis

Wout Hofman[1]
[1] TNO, Brassersplein 2,
Delft, The Netherlands
wout.hofman@tno.nl

Abstract. Current customs applications are declaration based to support the various customs procedures based on (inter)national laws and regulations. To be able to perform a proper supply chain risk analysis, customs requires to have all data in supply chains. The current declaration procedures are not sufficient since they do not supported retrieval of containers stuffing information resulting in the fact that authorities do not have a complete data set. It has been shown that enterprises already have a lot of data available to meet their customer requirements that can be made directly accessible to authorities, instead of 'pushing' data to the authorities based on procedures. By not only making this data available to customs but also to other authorities, they also comply with Single Window implementations. There are various solutions to data retrieval, e.g. a Service Oriented Architecture (SOA) offers a potential solution. The proposed approach in this paper is based on Linked Open Data (LOD) and implies innovative IT to be implemented by both authorities and enterprises in supply chains. The paper discusses LOD and its application to supply chain risk analysis. The proposed solution allows authorities to govern global supply chains in supply networks.

Keywords: Supply Chain Risk Analysis, Linked Open Data, Semantic Web, Ontology

1 Introduction

In a networked economy characterized by dynamic business relationships and of a global nature [1], trade volumes are rapidly growing. Globalisation and increased international trade are the two most important drivers for economic growth, which expose the population to new risks related to fraud, security, and safety [2]. In this context, the concept 'trusted trader' from a fiscal perspective was not only transformed to meet security requirements, but also supply chain security from a 'green lane' perspective is introduced. Whereas 'trusted trader' not only defines that a trader is known by authorities, but also has implemented particular compliance controls in its internal processes that can be audited by authorities. This concept is further extended by the EU FP6 funded ITAIDE project in I3 framework to construct a trusted trader network for the earlier mentioned 'green lanes' that are operated by trusted traders [3]. Information transparency or enterprise interoperability is one of

the important aspects of I3, not only between businesses (Business to Business: B2B), but also between business and government (B2G: Business to Government) [4]. Information transparency must offer authorities full supply chain visibility based on all available data relevant from the perspective of the physical process. Authorities like customs have defined various procedures in laws and regulations, but they still lack all data to get a complete view of supply chains. To meet these supply chain visibility requirements, authorities and traders can implement different technological solutions, e.g. in a declaration based approach, business documents are exchanged either on paper or using electronic formats like Electronic Data Interchange (EDI) and XML Schema (XML: eXtensible Markup Language). ITAIDE introduces a Service Oriented Architecture (SOA) [5] for supply chain visibility by authorities and other types of architectural approaches are also feasible, e.g. Event-Driven Architecture (EDA, [6]) or a combination of both (Event-Driven Service Oriented Architecture, EDSOA, [7]). All of these architectures are technological solutions for data capture by authorities. Data semantics is implicitly specified in this technology; extensions have been made to these syntaxes for explicit representation of semantics, e.g. Semantic Annotations for Web Services (SAWSDL [8]). Furthermore, each of these solutions requires additional specifications to be implemented by traders and authorities, potentially leading to an increase of the administrative burden. Instead of decoupling systems that leads to a decrease of administrative burden, introduction of the aforementioned technological solutions for full supply chain visibility leads to a tighter coupling of traders and authorities.

Linked Open Data (LOD) for data and content capture from its original resources is an innovative approach [9] that requires a minimal set of agreements to be implemented by traders and authorities, thus potentially decreasing the administrative burden and making optimal use of available supply chain data. LOD is an application of the so-called Semantic Web that foresees three types of applications [10], namely the crawling pattern, the on the fly deference pattern dynamically removing links for answering queries, and query federation by following links. The application of LOD and these patterns to supply chain visibility for risk analysis will be described in more detail in this paper and we will argue that the crawling pattern optimally supports supply chain visibility for risk analysis.

First of all, the objective of supply chain visibility for supply chain risk analysis and missing data is briefly presented and secondly the principles of Linked Open Data are presented. These principles are applied to supply chain visibility and finally conclusions and next steps are given.

2 Supply Chain Visibility for Supply Chain Risk Analysis

This section briefly presents the need of supply chain visibility for supply chain risk analysis. Furthermore, it defines the challenges for authorities for completing supply chain data with current missing data. Different solutions are presented in this section, whereas the next sections present a solution based on LOD in more detail.

2.1 The Need for Supply Chain Visibility

Various authorities like customs monitor value exchange crossing national or EU borders from a fiscal, security and safety perspective [2]. These authorities have agreed to monitor events based on (inter)national laws and regulations, which can lead to actual physical inspection. Examples of such events are selling products that may lead to export, buying leading to import, and containers with these packaged products leaving (exit) or entering a country (entry). There are particular regulations for intermediate storage, re-exportation, storage in bonded warehouses, etc. [11]. Basically, authorities currently have a requirement of receiving all information for those discrete events implemented by procedures; they themselves have to interrelate the events, i.e. to be sure that all exported goods also leave the country and all goods that leave the country have a relevant previous procedure (e.g. export, re-export), and they are able to perform risk analysis based on the captured data.

By monitoring discrete events, not all required data may be present, e.g. the export, exit, entry and import declarations do not contain packaging details of containers implying that the complete content of the container is not always to customs. To complete the information, the seamless integrated data pipeline is introduced consisting of all traders with their data and business documents participating in a supply chain (**Fig. 1**, [12]). As these traders share a lot of information for performing their business processes, it is the objective to re-use this information. It implies that visibility for authorities in supply chains needs to be increased.

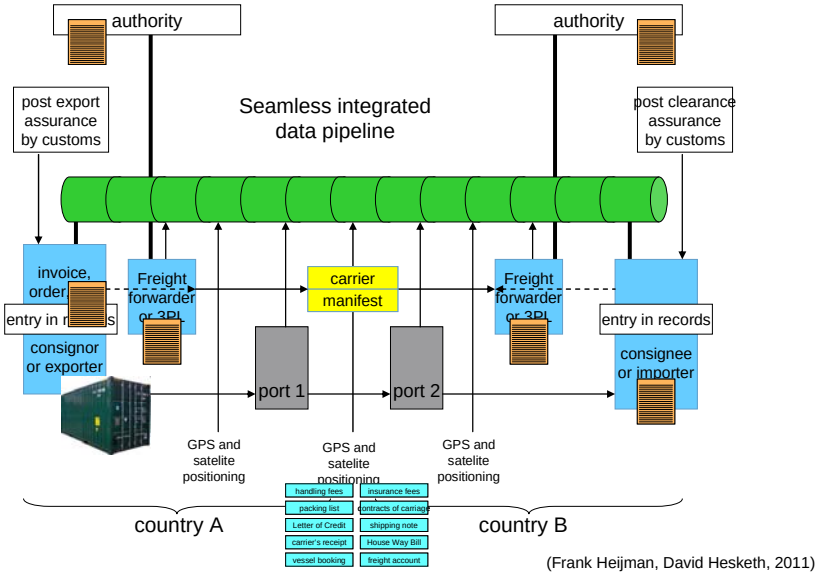

Fig. 1. Seamless integrated data pipeline

The conceptual pipeline can be viewed from two perspectives, namely a process and a data perspective. The process perspective specifies relations between traders in supply chains, e.g. a stevedore with contractual obligations to a shipping line and a

forwarder that arranges pre- or on-carriage to a port. These processes of cooperating traders can be described by transaction trees reflecting the business transactions between traders. The data perspective not only reflects the business transactions, but also the physical objects and their status. **Fig. 2** shows the data perspective representing the physical objects. Each of the relations between two high level data concepts can be created physically by another supply chain actor, e.g. stuffing containers can be done by a groupage centre and a stevedore performs loading and discharging. The physical objects can be more detailed and more physical objects can be added. 'Customs goods' is a particular view on physical objects. They represent physical objects in terms of a customs classification used for instance for VAT purposes, the so-called harmonised goods code. The status of physical objects is represented by their availability in a place, e.g. at a stevedores location. This availability can be provided by RF tags and business transactions amongst traders. These business transactions refer to a business activity that specifies conditions under which transactions can be performed.

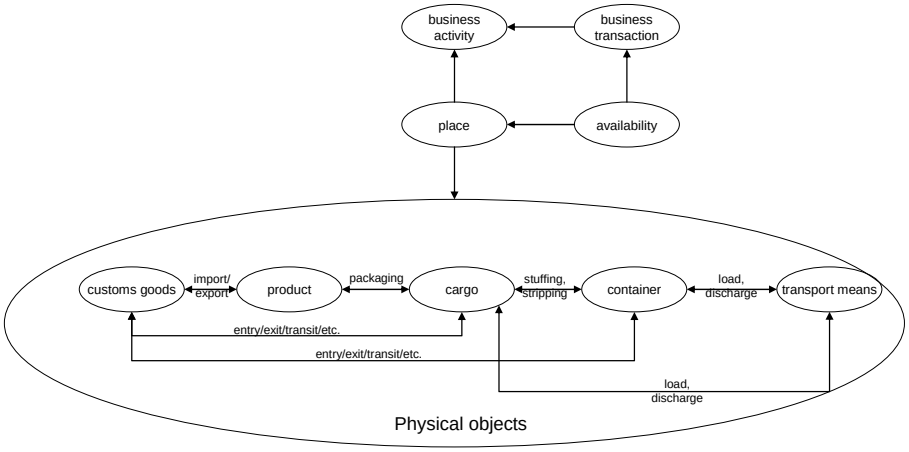

Fig. 2. Data perspective of the pipeline

In international container transport, authorities are not always aware of the 'packaging' and 'stuffing/stripping' relations. These are added by traders that are not obliged to perform a declaration. To retrieve this information, customs has several options that will be presented hereafter.

2.2 Options for Completing the Data Perspective

To complete the data perspective, customs basically has two options, namely monitoring more discrete events with accompanying declarations or a continuous monitoring of involved traders based on capturing their supply chain data (piggy backing, see [2]). Monitoring more discrete events can be supported by various technical solutions, e.g. messaging or an Event Driven Architecture combined with web services [7]. However, introduction of additional discrete events with customs

procedures increases the administrative burden for traders, meaning that they have to provide more data with new procedures. A continuous monitoring based on semantic web technology might offer an alternative that could even decrease the administrative burden whilst optimal making use of available data. This paper discuss that option.

3 Linked Open Data

This section explains Linked Open Data (LOD) in the context of the semantic web. An architecture for the semantic web defines data resources and data capture mechanisms according to a known semantics. These two aspects of the semantic web are described in this section.

3.1 Data Resources with Semantics and Metadata

Data semantics is the basis of the semantic web. Open standards for the semantic web are currently applied in many open data projects [10]. The approach enables the integration of many heterogeneous data in different sources by constructing links between that data. RDF is applied for documenting these links between 'subject' and 'object'. Subjects and objects are linked by their URIs (Uniform Resource Identifier). A typical example is that 'a person' (subject) 'has' (predicate) 'a name' (object). The combination of subject, predicate, object is called an RDF triple. RDF has limited functionality to specify semantics; this can be done with other open standards. OWL and other open standards like SKOS (Simple Knowledge Organization System) can be used for representation of semantics. Thus, a subject or an object in RDF can have complex semantics specified by an OWL or other document.

It is possible to distinguish various data resources. Sensors, enterprises, IT systems, social media are some examples of data resources. Data resources provide data of different natures, e.g. a sensor like an RF tag is a data resource with possibly streaming data and data fusion of this sensor data results in a new data resource. In all occasions, metadata has to be related to the data, specifying quality aspects of that data. Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH, [15]) and Dublin Core (DC, [16]) are two examples of metadata. Sensor Web Enablement [17] also specifies metadata for real life sensor information. Fig. **3** shows the relations between the aforementioned technologies. It distinguishes between open data with its metadata and links that are accessible via a URI, and the specification of semantics by ontology and metadata. The specification of semantics is also a data resource with a URI and thus is also considered to be open data. In ideal application, the data and content is directly accessible from a data store. A SPARQL (SPARQL is a specific query language for RDF) endpoint to that store could serve for direct querying the data. Most of the current applications based on open data require interpretation by end-users [10]. Semantics is required for scalability.

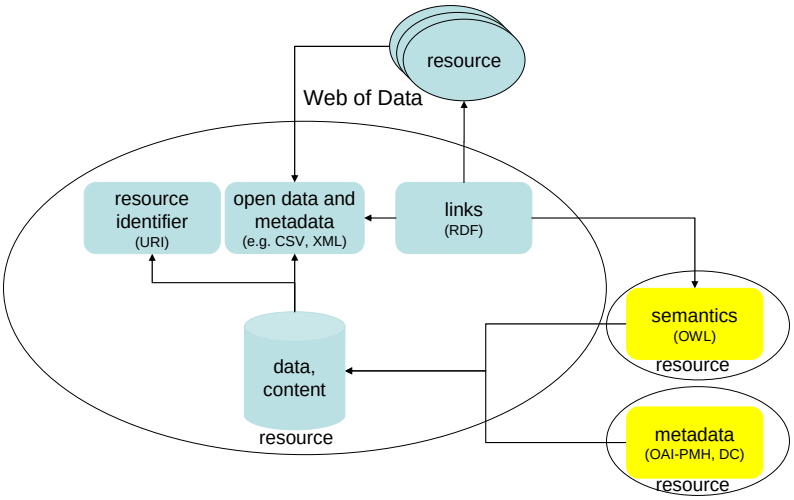

Fig. 3. Specification of a data resource

3.2 Data Capture from Data Resources

There are basically three ways to capture data from different resources [10], namely crawling, on the fly dereferencing for capturing data from resource chains and query federation. These will be discussed in more detail. Crawling data resources is based on retrieving all open data, metadata and links of those resources. Data capture and data analysis are separate functions in this pattern, decoupled by a data store (**Fig. 4**). Crawling constitutes a new resource with its own particular semantics and metadata that can be queried. A pipeline can be constructed for crawling each data resource. These pipelines can differ per resources, e.g. structured and unstructured data (content) can be crawled separately.

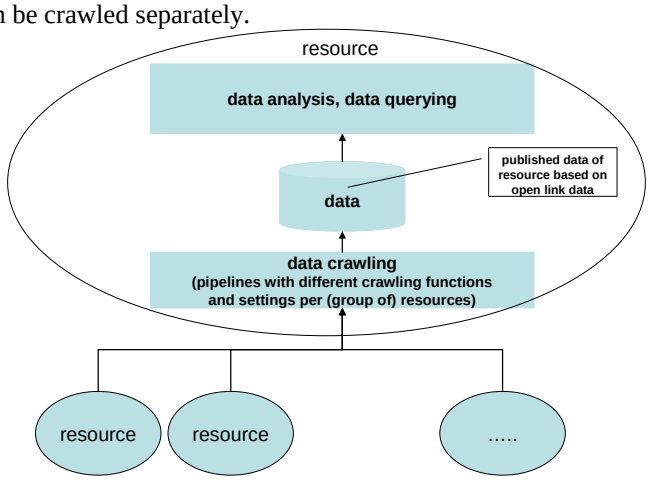

Fig. 4. Data crawling, indexing and data analysis/querying/fusion

Although crawling is able to capture data with different semantics, agreement on these semantics is required for analysis. Ontology can specify the structure of the data store. Possibly, data pipelines require transformation functionality for storing data in the data store according this ontology. For analysis purposes, additional metadata is required, e.g. the data resource, the way the data is captured by that resource, the time of data capture, and a reference to algorithms like SPSS (Statistical Package for the Social Sciences) used for data analysis and fusion performed by the resource. The metadata of the crawled resources also needs to be stored with the data. Data fusion can for instance be performed on (real time streaming) data of one or more sensors. The fused data has its particular metadata that has to be linked to the original resource data. Data fusion is not only applicable for real time data streams, but can be applied to all types of data. In such a way, aggregated and processed data is created. Crawling can thus be applied for analysis of large amounts of data and requires replication of that data.

The second way of data capture is the so-called on the fly dereferencing pattern implying that all data is captured by following links between resources. This pattern captures data from a resource chain based on links between those resources. Only that data is captured that is relevant for the resource chain. The data resources from which data is captured are not known in advance, but become apparent by following links. Supply chains can be seen as an example of resource chains.

The third pattern is called the query federation pattern. This latter pattern is based on sending complex queries directly to a predefined set of data sources. This pattern can be used for one time queries for which the data resources to be queried are known.

4 Supply chain visibility with Linked Open Data

As we have indicated, Linked Open Data is the most commonly known application of the semantic web. This section constructs views a trader as a data resource and proposes a means for data capture to create supply chain visibility to customs. Firstly, a proposal for data capture is presented; secondly this section gives the conditions for implementing this proposal, and finally, advantages of the proposed solution are discussed.

4.1 Data Capture by Piggy Backing on Supply Chain Data

The options crawling and on the fly dereferencing seem to be the most applicable implementation options for supply chain visibility. Whereas on the fly dereferencing dynamically constructs individual supply chains in a logistics value web, crawling captures data from all actors in a value web. On the fly dereferencing is based on links between traders. Business transactions (see before) can serve as links, but only business transaction related to particular physical objects like containers need to be evaluated to construct a supply chain. Constructing supply chains in this way may be time consuming and only that part of the supply chain is constructed for which the

links can be followed at a given time. As many traders will act in more than one supply chain, on the fly dereferencing may not be the best option.

Thus, crawling seems to be a better option for data capture by customs. Crawling decouples data fusion and analysis functionality from data capture, allowing performing data analysis independent of data resource availability. It means that supply chain data of traders and links between those traders based on business transactions are captured independent of supply chain risk analysis. Each trader acts as a data resource in a value web by publishing available supply chain data (piggy backing). A link must have meta-data like a validity period to be able to distinguish between operational and historic transactions. The business transaction links can be used to reconstruct supply chains in the value web. Authorities can decide themselves how frequent they capture data this way; it can be every 5 minutes but also on a daily basis. Data capture also depends on the availability of data resources. Agreements have to be made in this respect between authorities and companies. Supply chain data can be enhanced for analysis purposes by retrieving additional data from external source, e.g. not only databases with know traders of authorities that can be trusted, but also data from social media defining relations between persons and thus companies. Social media data is not always trustworthy and needs to be handled as such.

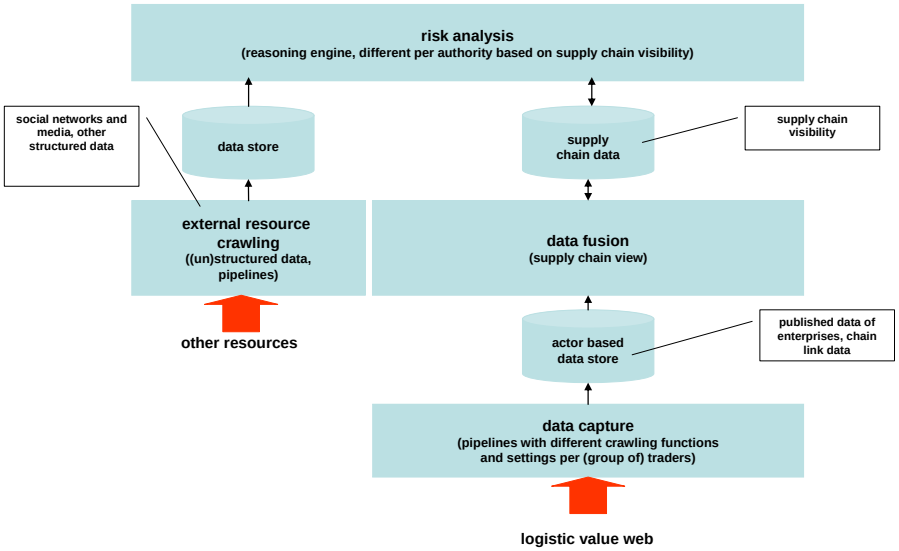

Fig. 5. Supply chain risk analysis based on capturing data in value webs

Supply chain and business transaction data (**Fig. 2**) published by traders need to be refreshed based on agreed events, meaning that authorities still need to monitor progress of supply chains in accordance with laws and regulations. These events can be the ones that are already defined by current customs procedures, e.g. exit, entry, import, and (re-)export (see before). A trader acting as a data resource has to build in these controls for refreshing the data, but basically they will be available since traders do business with each other.

An authority has two options for risk analysis after data capture. The first option is to store all data as received and analyse this data. It means that data is duplicated based on crawling frequencies. It may lead to potential large data stores, depending on the number of logistics movements that has to be captured. The second option is to fuse the received data to actually reflect the structure of a supply chain, i.e. container data is only stored once for one supply chain thus creating supply chain visibility (**Fig. 5**). The latter situation can be constructed by following links between actors based on transactions. The proposed solution shown in **Fig. 5** can be optimized, e.g. to capture only data that is refreshed by a trader or to capture only transaction data of traders, construct supply chains in a network, and analyze vulnerability of these chains based on additional data of traders. In case one of these sources is considered to be a threat, additional data can be captured for further analysis. The latter approach closely relates to what is called System Based Auditing [14].

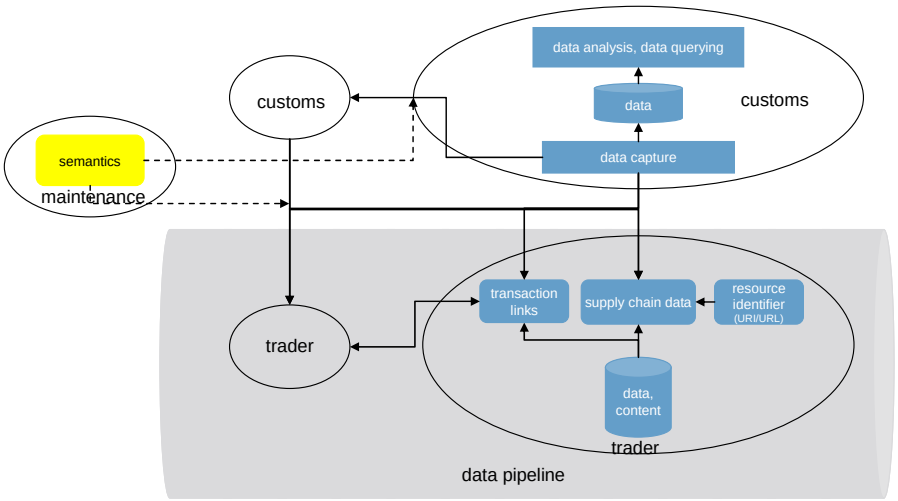

Fig. 6. Global supply risk analysis

The supply chain data store offering supply chain visibility is a new data resource that can be accessed by other resources. Supply chain data is the basis for various other applications, e.g. economic figures, statistics, different risk analysis functions for each authority, etc. By further enhancing supply chain data with results of risk analysis, supply chain data of one (customs) authority can be used by another authority thus constituting a global network of interconnected authorities. Chain data crawling thus does not only mean capturing data of supply chain traders, but also data of authorities in other countries, possibly with the inclusion of the risk analysis results of that other authority. The latter requires a level of trust amongst authorities in different countries. **Fig. 6** shows that a supply chain pipeline (**Fig. 1**) can thus be monitored by several authorities, each from its own perspective and with its particular responsibility. Global supply networks can thus be monitored by more that one authority.

4.2 Ontology as a Basic Condition

As we have stated, semantics of open data is one of the main conditions for the semantic web and thus for global supply risk analysis based on supply chain visibility (**Fig. 6**). Semantics can be represented in different ways. In the semantic web, a data resource can also contain semantics of other data resources. Such a reference is feasible if one of the semantic web standards is used, for instance Web Ontology Language (OWL). Currently, the World Customs Organization (WCO) has specified semantics for declarations supported by messaging with a UML class diagram covering all types of laws and regulations for global logistics [11]. There are two issues relevant in this context namely re-use of concepts and definitions of this class diagram and the functionality of the WCO class diagram. With respect to re-use, different concepts of the data model can currently only be copied and not referred to via an URI. Representing such a class diagram as ontology makes it accessible for all authorities and supply chain enterprises and allows them to construct IT based on ontology, without re-keying definitions, etc. Furthermore, applying the concept 'networked ontology' [13] makes it possible to construct dependencies between different ontologies. Complete ontologies can be imported and for instance concepts of these ontologies can be tied to equivalent concepts by the construct 'equivalentclasses' in OWL2 thus providing a means for matching and re-use of existing concepts.

The second aspect is the functionality supported by the WCO class diagram. It has been specified to support all data relevant for governing supply chains by authorities and constitutes not only customs specific data like harmonized goods code, but also container and vessel data. One of the basic questions is whether this data is sufficient to support all processes of actors in supply chains. It is most likely that a networked ontology for open data in supply chains needs to be constructed based on a representation of the WCO class diagram by an ontology. Furthermore, localizations are most probably required, e.g. a national authority may have additional data requirements. By constructing networked ontologies, localizations can easily be supported.

4.3 Advantages

Implementing the proposed solution has many advantages for both traders and authorities. Whilst the proposed way of data capturing is a complete decoupling between traders and authorities with a minimal set of agreements, it will decrease the administrative burden. Traders, being shippers and logistic service providers, publish their supply chain data according to an agreed ontology including the transaction links. By making their data available to authorities, they can adhere to requirements of all laws and regulations, independent of the way they are implemented. There are lots of authority initiatives to change the implementation for lessening the administrative burden for traders [14]. By implementing the proposed approach, traders and logistic service providers can adhere to all initiatives.

Authorities on the other hand can optimize the physical inspection of supply chains, because all supply chains in a logistic value web are completely visible.

Authorities do not have to introduce additional events that have to be monitored by new procedures supported with messaging, but have full visibility if all traders make their data available to those authorities. Each chain but also each individual trader can be analysed on its behaviour. In case traders have data available, packaging and stuffing data will also be available to authorities (**Fig. 2**). Supply chain visibility thus improves the detection of anomalies by also including external data resources in the analysis. Furthermore, each authority can have its specific risk analysis method; for the purpose of a seamless goods flow they need to align their inspection planning. The latter prevents that the same goods are inspected more than once by different authorities. The proposed supply chain risk analysis pattern requires a new approach to IT of those authorities. They need to capture data from all types of resources and use this data for analysis instead of keeping validating the completeness of a data administration. Information management will change.

5 Conclusions and Next Steps

This paper takes a semantic web view for supply chain visibility and proposes a solution for data capture in logistic value webs. Supply chain visibility will improve supply chain risk analysis and the proposed data capturing mechanism will decrease the administrative burden since it is based on already available supply chain data (piggy backing), whereas other solutions require the implementation of new procedures by traders and authorities that will increase the administrative burden. Semantics of supply chain data can be specified by an ontology that can be based on the WCO class diagram.

It is not required for authorities and traders to implement the proposed solution instead of current, existing declaration based solutions. These can still be used and are also a data resource for data capture to construct supply chain visibility. However, as indicated in section 2, the current declaration based systems do not offer complete supply chain visibility. Thus, additional software is provided offering visibility of those events that are currently not declared. Commercial supply chain visibility tools based on sensor (RF) data like EPCIS (Electronic Product Code Information Services) or SICIS (Shared Intermodal Container Information System, www.integrity-supplychain.eu) could be used for these purposes. Not only adoption, technical and organizational aspects are of further study, but the solution also has to fit in current laws and regulations. Furthermore, the business case for both traders and authorities has to be made as part of the adoption. Security and privacy also have to be solved based on policies of traders and authorities.

Acknowledgement

This paper is the result of the Dutch Extended Single Window project funded by Dinalog and the cooperation between TNO and Capgemini for developing innovative approaches for supply chain visibility. These concepts are further developed and tested in EU FP7 Cassandra (www.cassandra-project.eu).

References

1. Heineke J., & Davis M. (2007), The emergence of service operations management as an academic discipline, *Journal of Operations Management 25 (2007)* 364–374.
2. Rukanova B., Bjørn-Andersen N., Ipenburg F. van, Klein S., Smit G., & Tan Y.-H. (2011), Introduction, in Accelerating Global Supply Chains with IT-innovations, Tan Y.H. et al. (eds.), Springer.
3. Henningsson S., Budel R., Gal U., & Tan Y.-H. (2011), ITAIDE Information Infrastructure (I3) Framework, in Accelerating Global Supply Chains with IT-innovations, Tan Y.H. et al. (eds.), Springer.
4. European Interoperability Framework for Public Services (EIF, (2009), European Commission – IADBC.
5. Erl T. (2005), Service-Oriented Architecture – concepts, technology, and design, Prentice Hall.
6. Hophe G, & Woolf B. (2004), Enterprise Integration Patterns- designing, building, and deploying messaging solutions, Addison-Wesley.
7. Overbeek S., Klievink B., & Janssen M. (2009), A flexible, event-driven, Service-Oriented Architecture for orchestrating service delivery, IEEE Intelligent Systems.
8. Semantic Annotations for WSDL and XML Schema (2008), W3C Recommendation.
9. Berners-Lee T., Hendler J, & Lassila O. (2001), The Semantic Web, a new form of web content that is meaningful to computers will unleash a revolution of new possibilities, Scientific American.
10. Heath T., & Bizer C. (2011), *Linked Data – evolving the Web into a Global Data Space*, Synthesis Lectures on the Semantic web: Theory and Technology, Morgan & Claypool Publishers.
11. World Customs Organization (2009), WCO Datamodel – cross border transactions on the fast track.
12. Heshket D. (2010), Weakness in the supply chain: who packed the box?, World Customs Journal, volume 4, number 2.
13. Haase P (2009), D1.1.5 – Updated version of the networked ontology model, EU F6 NeOn: Lifecycle support for networked ontologies.
14. Stijn E. van, Phuaphanthong T., Keretho S., Pikart M., Hofman W., Tan Y-H. (2011), Implementation Framework for e-solutions for Trade Facilitation, Tan Y-H., Bjørn-Andersen N., Klein S., Rukanova B. (editors), Accelerating Global Supply Chains with IT-innovations, Springer.
15. Open Archives Initiative – Protocol for Metadata Harvesting, Protocol Version 2.0 of 2002-06-14, Document Version 2008-12-07T20:42:00Z, http://www.openarchives.org/ OAI/ 2.0/ openarchivesprotocol.htm.
16. Dublin Core Metadata Initiative, DCMI Metadata terms, http://dublincore.org/ specifications.
17. Botts M., Percivall G., Reed C. and Davidson J. (2007), OGC Sensor Web Enablement: overview and high level architecture, OGC White Paper.
18. EPC Information Services (EPCIS) Version 1.0.1 Specification (2007), GS1.