Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1203–1209

Applying Big Data and Linked Data Concepts in Supply Chains Management

Silva Robak
Uniwersytet Zielonogórski, ul. prof. Z. Szafrana 4a, 65-516 Zielona Góra, Poland
Email: s.robak@wmie.uz.zgora.pl

Bogdan Franczyk
Uniwersytet Ekonomiczny we Wrocławiu, ul. Komandorska 118/120, 53-345 Wroclaw, Universität Leipzig, Germany
Email: franczyk@wifa.uni-leipzig.de, bogdan.franczyk@ue.wroc.pl

Marcin Robak
XLogics Sp. z o.o., ul. Kostrzyńska 4, 65-127 Zielona Góra, Poland
Email: m.robak@xlogics.eu, Uniwersytet Zielonogórski, WEIiT, m.robak@weit.uz.zgora.pl

Abstract—**One of the contemporary problems, and at the same time a big opportunity, in business networks of supply chains are the issues associated with the vast amounts of data arising there. The data may be utilized by the decision support systems in logistics; nevertheless, often there is an information integration problem. The problems with information interchange are related to issues with exchange between independently designed data systems. The networked supply chains will need appropriate IT architectures to support the cooperating business units utilizing structured and unstructured big data and the mechanisms to integrate data in heterogeneous supply chains. In this paper we analyze the capabilities of the big data technology architectures with cloud computing under usage of Linked Data in business process management in supply chains to cope with unstructured near-time data and data silos problems. We present our approach on a 4PL (Fourth-party Logistics) integrator business process example.**

I. Introduction

In the contemporary world the business companies have to face unprecedented challenges. As a result of globalization the amount of data arising in supply chains is raising, the competition is becoming fiercer and the customers often expect integrated services, what requires a close cooperation between several involved organizations. The companies have to adapt to new, such as networked, business models and rethink their role and position in their value chain regarding the potential possibilities given by the utilization of big data to add value for their customer and suppliers. This requires some changes from companies in their organizational view, but at the same time in their information technology view. The appropriate technology environment is needed to support their interoperable business cooperation.

The problem of the appropriate information technology environments for collaborative processes between business participants is twofold. Firstly, the appropriate IT infrastructure for utilization of big data is needed, and secondly, there are data 'siloes' from diverse applications. The last problem has been approached with several solutions like common IT platforms consisting of (possibly common) components based on established standards, standard enterprise information systems, and standard business protocols. Nevertheless, the IT environment platforms still contain proprietary applications like enterprise resource planning systems (ERP), customer relationship management systems (CRM), etc. The foreseen scale of collaboration between business partners may require undertaking further steps for IT environment integration, such as one of the known enterprise application integration solutions or usage of the Web services [1].

The inter-company networks are defined as complex arrays of relationships between companies, which establish these relationships by interacting with each other [2]. Whilst the markets are expanding toward inter-company networks (webs) of collaborating organizations, mentioned above IT integration solution approaches, do not seem to be sufficing and satisfactory. This level of an organizational form of the market participants requires mutual adjustment in information sharing and data management, and further a coordination of collaborative business processes of the supply chain's participants.

In the paper we will approach the problem of possibly advantageous utilization of vast amounts of data (in variety of formats) in supply chains and also the information integration issues in order to overcome the data silo problem. We will investigate the appropriate IT architectures for big data used in association of cloud computing facilities [3] and the utilization of common (open stated) data format as it is offered by Linked Data [4] for data silos integration purposes. We consider the network from a supply chain perspective with emphasis on the value-adding partnerships. The proposal of possible utilization of the Linked Data as an integration solution for business process management BPM in supply chains networks we have already presented in [5]. In this paper we investigate the usage of big data IT architectures, appropriate for supply chains in conjunction with data silos integration possibilities for supply chains on the basis of (open) Linked Data.

A supply chain is defined as a network that comprehends all the organizations and activities associated with the flow and transformation of goods, starting from raw material stage through the whole process, to the end user, as well as the associated information flow [6]. In the paper we will concentrate on the networked supply chain activities and information flow.

In the inter-organizational information systems, which link companies to their suppliers, distributors and customers, a movement of information through electronic links (e.g. XML/EDI - Extensible Markup Language/ Electronic Data Interchange) takes place across organizational boundaries between separately owned organizations. It requires not only electronic linkage in form of basic electronic data interchange systems (as for purchase orders, delivery notes, cash flows, etc.), but also interactions between complex cash management systems or by accessing shared technical databases. So the problems with sharing and exchange of information are still viable in supply chains contexts.

The existing EDI standard [7], as message-centric solution, has limited possibilities in enabling interactions in the value chain. The representation of business processes and vocabularies in a domain to potentially automate the trading partners interactions is missing. Another important aspect regarding supply chains networks is integration of additional data from semantic Web applications into logistic systems.

A business process consists of one or more than one related activities that combined together respond to the need for a business action [6]. The processing steps in a workflow might go through numerous data transformations (geographic, technological, linguistic, syntactical and semantic transformations). Communication is an important part of the process and (e-) business processes exist within certain environments. In the dynamic business environment, such as networks of venture participants involved in logistic value chains, where coordination problems in the business process management plays the key role, the appropriate IT architecture, data amount and format, are essential.

Therefore, as stated previously in our paper, we will analyze big data architectures and Linked Data for business processes in supply chains in business networks. For this aim the rest of the paper is organized as follows.

In Section 2 we characterize main features of big data and architectural elements needed by IT infrastructures to support big data during business process management in supply chains. In Section 3 we summarize the Linked Data principles and concepts. In Section 4 we provide an example scenario for 4PL (Fourth-party Logistics) integrator managing the package delivery from Webshops in Asia to customers in Europe. We examine the possible added value resulting from usage of big data and open Linked Data elements that may possibly be useful for the purpose of decision support in supply chain networks. We will also try to show (in Section 5) how the information integration on the base of the Linked Data in conjunction with the big data IT architectures may be applied to achieve the improvement of supply chain environments. In the last Section we conclude our work.

II. Big Data

A. Big Data Features

The big scale usage of available and generated data is made possible for organizations owing to cloud computing paradigms, such as Infrastructure as a Service (IaaS), Storage as a Service (SaaS), which revolutionized the way the computing infrastructures are used [3]. Big data is referred to data that goes beyond the processing capacity of the conventional database systems. In addition to the aspect that it is big (e.g. a huge number of small transactions, or continuous data streams from sensors, mobile devices etc.) it may move too fast, or does not fit the structure of traditional (i.e. relational) database architectures. Big data also may have a low value for further usage before processing it [8].

According to [8] when we denote a big amount of data as "big data" it has to cover the three "Vs" (features) such as: volume, velocity and variety. Other authors (e.g. [9], [10]) add the fourth V-feature: value.

The first feature - volume of big data - denotes its massive character. The big volume of data is beneficial for the data analysts. It may improve the analytics models by having more cases available for forecasts and increase the number of factors to be considered in the models making them more accurate. Nevertheless, the volume bears potential challenge for IT infrastructures to deal with big amounts of data, especially when taking into account its second feature – velocity.

The second feature of big data is the velocity in which data flows into organization or the expected response time to the data. Big data may arrive quickly - in real-time, or near real-time (denoted in this paper as near-time). If data arrives too quickly the IT infrastructures of the organization may be not able to respond timely to it, or even to store all of it. Such situations may lead to data inconsistencies. We will regard further the issue of possible velocity consequences in the next section considering the suitable architectures for big data applications.

The third feature of big data is the variety of data. Big data may have diverse structures and forms, not falling into the rigid relational structures of SQL databases without loss of information. Some of data may be saved as blobs in inside traditional data bases. Therefore the IT infrastructures for big data are denoted as NoSQL, which means that data is "not only SQL" [10]. Several examples for diverse kinds of data are standard business documents, transactional records, and unstructured data in form of images, recordings, HTML documents (web pages), text and email messages, streams from meters and environment sensors, GPS tracks, click streams from Web queries, social media updates, data streams from machines' communication or wearable computing sensors, and many others.

The big data value feature denotes the need for processing it before using it in order to make it valuable for analysis purposes.

B. Big Data Architectures

In the previous section the four characteristic features of big data have been discussed. It is apparent that conventional IT structures may encounter problems with storing variety of data and immediately reacting to it. Firstly, it is because of big data amounts on unstructured data arriving in near-time. The fact that data is unstructured, or rather, it lacks a structure appropriate for storage in conventional SQL databases, implies that other solutions will be needed.

The first issue to consider is the common usage of SQL data bases. IT infrastructures in supply chains include structured data in form of OLTP (Online Transaction Processing) and OLAP (Online Analytical Processing) systems. While the traditional OLTP systems support the transactional systems with highly structured SQL databases, the OLAP systems contain aggregated historical data in form of cubes. The OLTP systems deliver simple reports, while OLAP systems (known as Data Warehouses) are suited for (traditional) business intelligence applications with reporting facilities on business statistics, performance, etc. on the basis of structured (analytical) historical data. These both databases forms are unsuitable for big data purposes. The data stored there has to have fixed structure, which is conflicting with variety of big data. The OLAP and OLTP contain only high quality data, what is not the situation in case of low value big data (on the opposite to low value big data).

Another problem arises due to the velocity of big data. For the reason that analytical OLAP systems contain only historical data, they are unsuitable for big data applications.

The rigid SQL data structures are insufficient for big data applications, but there are other OLT solutions like “NoSQL OLTP” – MongoDB, AmazonDynamo or Windows Azure Table Storage [10]. This type of database is known as the ‘key-value stores’ where the data is stored by key and its value is a blob and this solution is widely adopted by enterprises [3].

There is also a known solution for “NoSQL warehousing” for storing and analyzing massive data sets – Apache Hadoop [11]. The Hadoop is a framework for development of open-source software with its own highly distributed HDFS file system, MapReduce framework for writing and executing distributed algorithms and its own query languages – Hive and Pig. The Hadoop components are not only highly distributed but also high tolerant.

Another aspect of big data which is different from SQL databases, is that the results from big data analysis are immediately used and often discarded after that. If not, some bridging solutions are needed, e.g. SQQOP for connecting SQL and Hadoop [10], but they may turn inefficient. Moreover, for handling variety of big data, another solutions like dedicated XML store or graph databases are available [12].

The volume feature of big data can be handled with the usage of capacities and platforms offered by cloud computing [3]. With the pay-as-you-go and low time-to-market solutions, it became affordable even for small organizations.

The big data applications are possible as combinations of diverse technologies (products mash-ups) [12].

The volume, variety and value problems of big data can be tackled by solutions mentioned above. There still remains the dilemma with big data velocity. If data streams arrive quickly in real-time (or near-time) there may be a problem with storing them. One solution possibility would be temporary batch of data in the data pools. This problem is resolved by the Lambda architecture proposed in [13]. Its authors assume that a query is a function on the whole data pool:

```
query = function (all data)
```

Therefore Lambda is a three layers architecture with the batch layer, serving layer and speed layer (see Fig. 1).

The bottom layer (the batch layer) is dedicated to the precomputing on the whole data pool. Its two functions are storing master data set and computation of arbitrary (i.e. any) views. The batch processing principle is well known; for big data the Hadoop is a canonical example [13]. In the Lambda architecture the batch layer continuously recomputes the batch view from scratch.

The middle layer, the serving layer delivers the random accesss to batch views and also is updated by the batch layer.

The top layer, the speed layer deals only with the latest data received during running precomputations in batch layers. According to [13], it compensates for high latency of updates to the serving layer and uses fast incremental algorithms; the batch layer ultimately overrides the speed layer.

In [13], there are big data applications examples as combinations of diverse technologies (products mash-ups) and providers of big data solutions, such as SAP, Oracle, IBM, HP, etc.

With the usage of Hadoop a very fast historical data analysis will be possible based on the Hadoop file system and the MapReduce technology. Nevertheless, it would not be sufficient for some kinds of applications, because the data analysis of current incoming data would be missing. Therefore, in addition to Hadoop we may use the Complex event processing technology (CEP) [14] for dealing with the huge amounts of other real-time data incoming during the running processes. With data supplied to the system and with the appropriate rule sets a dedicated decision support system would be able to react in a suitable way in critical situations, each time when a near-real-time decision will be needed.

The main idea of the introduced concept includes two aspects: the historical context of the Hadoop data, while reacting to the current situation with the CEP technology

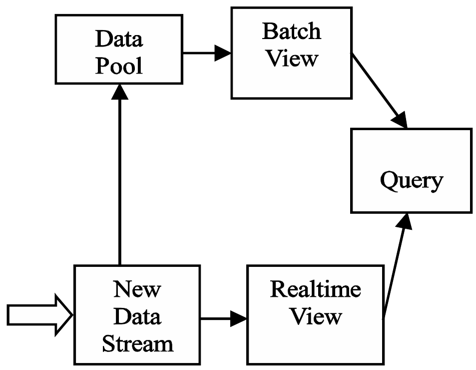

Fig. 1 Lambda architecture diagram [12]

usage. Merging of actual and historical data would add a new value in the decision making processes.

III. Linked (Ppen) Data

The Linked Data principles were introduced by Tim Berners-Lee at the TED 2009 presentation [15]. He outlined four rules for making human or machine-readable links for the exploration of web of data. The first rule refers to the usage of Uniform Resource Identifier URI [16] for identification of items (called “things”). The second given rule specifies that only HTTP URIs are meant, so that people can look at them and these can be found by the standard established Domain Name Space (DNS) system. The third rule was formulated for the purpose of providing additional useful information for the items defined by URI. The information should be denoted in a standard format, such as Resource Description Framework RDF* [17], in form of RDF/XML or an alternative serialization (N3, Turtle). The last, fourth rule concerns providing the linkage of such described items with other related items (data), so that the related information on the Web can be discovered more easily.

The further development was the Linked *Open* Data LOD, the concept recommended by the World Wide Web Consortium W3C [4]. It is a star rating system of linked data that allows for proving to which extent the linked data can be regarded as open. The rating is formulated as a five principles scheme, where each next scheme principle extends (for the next star added) the former one by integrating an additional feature. The first principle states that the data should be available on the Web, no matter in which format, but it should be one with an open license. The second principle adds that it should be machine-readable structured data. The third principle adds that the format of the data should be a non-proprietary format. The fourth principle assumes achievement of the former three principles and additionally presumes the usage of an open W3C standard for identification of items, like RDF (RDF/XML, N3 or Turtle) or SPARQL [4] formats (SPARQL Protocol and RDF Query Language) for larger data amounts of data sets. The last principle, required for getting the five star grading assumes contextual linkage of rated data to other resources described in the same way.

IV. 4PL Integrator Example

The development in contemporary logistic networks leans toward possible outsourcing of various logistic functions or services. Further trends include possible integration of outsourced functions/services or even the outsourcing of the whole business processes. At present the dominant are so-called 3PL (Third-party Logistics) solutions [18]. At the next developmental stage the concept of the 4PL (Fourth-party Logistics) emerged. It encompasses the functions offered by a 4PL logistic provider, which is acting as an integrator, assembling the resources, capabilities and technology needed for design, building and running of the comprehensive supply chain solutions [19]. The international 4PL do not need to have their own transport [20]. They may work directly with companies offering transport, or with the 3PL providers, what includes different kinds of carriers, consolidators and forwarders such as ocean carriers, airfreight forwarders and local carriers. The 4PL govern the settlements of the agreements with all involved partners.

As an example of the ideas presented in previous Sections, we consider an example of a 4PL logistic provider which is managing the shipping of commodities bought from the Webshops located in Asia by the customers residing in Europe. The Webshops in our example are located in different Asian cities and offer toys and consumer electronics goods. Our 4PL logistic integrator outsources a warehouse (a hub) in Asia, so that he can consolidate the shipments, which are sent from different Webshops for their further transport to Europe. He also outsources space and commissioning capacity at few hubs in Europe (i.e., near London and Lyon), where the goods first arrive from Asia. The 4PL integrator operates a software platform that integrates the orders from diverse shops and processes the communication with the integrated shipping software dedicated for labeling the shipments for the European carriers. It includes up-to-datedness of the solutions (i.e., carrier-dedicated label layouts) for the European market. Thus, the carrier integration purpose of the 4PL-software fulfills one of most important roles of the 4PL.

Below we will show a standard solution, which is traditionally offered by the 4PL and in a next Section we describe an improvement bringing added value, which could be potentially achieved by the supplementary usage of big data and open Linked Data.

In the basic scenario the 4PL relies on the Webshop`s order data and on the agreements with the freight forwarders and the carriers. At first the goods are ordered by the customers in Europe, who choose a particular European carrier company while ordering products. We consider the situation of delivering of valuable, bulky goods equipped with RFID. The ordered goods are then labeled with the European carrier shipping labels by the Webshops, which download these outputs from the 4PL's IT platform. This results in every single parcel having a shipping label, which fully complies with particular European carrier labeling specification and is augmented with the corresponding Webshop logo.

In the next step the labeled shipments from a particular Webshop are consolidated on palette(s) and brought per freight forwarders to the Asian hub, where the goods are re-consolidated (individual shipments from different Webshops are packed on palettes for a particular hub in Europe) and sent overseas to appropriate European hubs. In Europe the palettes are unpacked and the individual shipments are scanned as ready for a European carrier pickup. The carrier, which was selected by the Webshop customer, will transport the goods within Europe, after they have reached European hub. The carrier's driver receives the printed list with information about the parcels he takes from the hub. At the same time electronic information with the data of the shipment is sent via EDI to the carrier system. From this moment on tracking in Europe is possible, but on the carrier website only.

A. Enhanced 4PL Scenario

The enhanced 4PL's platform scenario performs the same tasks as in the above basic scenario, but the 4PL also gathers a lot of additional data, which will be used for the improvement of the transport decisions through the route.

Every individual shipment prepared by the Webshop is tracked in the 4PL's platform, starting from the point it leaves the Webshop and is picked by the Asian freight forwarder. The freight forwarder provides GPS tracking for the road carried pallet. Further scans are made in the Asian and in the European hubs. The tracking after this point is processed through the carrier tracking system; the corresponding data is imported into the 4PL's IT platform through web service requests, sftp status file transfers or by automated read outs of the carrier tracking website - depending on the solutions provided by the particular carrier company.

Import and analysis of tracking data from different carriers is a demanding task, which may be supported by Linked Data. Over 17 thousand status event descriptions, which are used by European carriers, can be synthesized to less than one hundred events. Thanks to the full tracking transparency the 4PL is able to collect and analyze the data, and deduct how long it takes to transport goods from point A to point B. In case of delayed shipments, which are reported in his platform in real-time, the 4PL has the possibility of picking out an additional feature such as an express route for further parcel transportation. The gathered tracking data also enables finding out the reasons for the delays, which could help avoiding these in the future.

Another important big data source gathered and analyzed by the 4PL's platform comprise the Asian weather reports and also road and airport traffic reports (available as open Linked Data). Such information is inevitable to support real-time decisions of choosing appropriate transport way (air, road) within Asia. For example, the road transport may turn faster, if the nearest airport is expected to be covered in fog for the next two days.

The 4PL may also use social media and blog data, so that the trends in popularity of e.g. the toys and electronics offered by the Webshops can be regularly observed and evaluated for different regions of Europe. Based on this information, order and transport volumes can be better forecasted, enabling preparation of appropriate transport routes (changing the agreements with the freight forwarders, carriers and warehouses/hubs) in case of forecasted booming or collapsing demand.

Among others tasks, the 4PL has to manage big amounts of data coming from the carriers and other participants in the SCM with different formats and further with diverse semantic interpretation for each identifier. For instance, each carrier has its unique scope of the services (in addition, not always available for all cases) with its own sets of identifiers and furthermore, of the possible package statuses.

Thus, there is a lot data mainly associated with the delivery status of the parcel, denoted in proprietary format. For instance the package, which has been delivered to the client, can get the status "delivered", "closed", "ready", etc. dependent on the carrier firm, etc. So the Integrator has to perform the task of mapping all the unique package status names to one standard format in order to be able to further process the associated data. This way also all the data with the same semantics gets reduced to one uniform internal name and format of the status. Broadly speaking, a company dealing with about hundred carriers in different countries has to understand about 20 thousand possible package statuses, which can be reduced by the 4PL integrator to about one hundred mapped status descriptions. These will be further needed in the supply chain EDI data exchange. The usage of the Linked Data could facilitate the mapping of the equivalent data, not only on the 4PL side, but also among other supply chain participants.

In the next Section we show the possibility of integration of logistic data by using open Linked Data facilities.

V. Data Integration With Linked Data In Value Chains

As stated at the previous Sections, the process steps in a workflow could undertake numerous transformations of data. A common format could improve the communication between the participants collaborating in the process environments and serve as a broker between SQL and NoSQL data, especially in the big data environments.

In our example, the source data acquired from a Webshop is, until delivery of the commodities to the customer, administered in the subsequent stages, changing format and being adjusted and enriched through numerous additional transformations, which are needed for accomplishment of the activities of the participants in a joint business venture.

The Semantic Web concept supports the basic idea of the Web considered as an open community sharing information around the world. As pointed out in our hypothetic Webshop parcels example, one part of the data integrated into the data exchange flow of business networks could be the data supplied into the 4PL integrator software platform from Web applications like the Geo, metrological or traffic data, partially enriched with semantic information described with RDF triples.

Since there still are no known established business solutions successfully working on the base of ontologies we consider application of Linked Data concepts as a more of a lightweight solution than the semantic description of data which is exchanged in networks connected through Internet and enriched with data from web applications like OpenStreetMap [22] or DBpedia [23].

The information assumed is to be presented as Linked Open Data, presumes that the data not in a proprietary format. Therefore it is important that the communication software (e.g. of 4PL) will support open data formats, such as CSV (Comma-separated values) [24] or transformation to such an open format.

The data exchanged between supply chain participants may be enriched with semantic information by means of the RDF graphs. At present time information is stored mostly in relational databases. There are some solutions for data transformations, i.e. Triplify for transforming of the data stored in transactional SQL databases into RDF representations [25]. Other possibilities, like object serialization or hierarchical representation, should be mapped into the graph data models. Meanwhile there are multiple semantic database implementations known, such as Triplestores, a purpose-built database type dedicated for the storage and retrieval of triples, e.g. Virtuoso [26]. In addition to queries the triples can be imported and exported using RDF or other formats.

The mapping between customized IT solutions and different data formats into the triple representation could be undertaken by means of dedicated software.

The usage of open formats with RDF-defined semantic could support easier data entries into the digital value chains. The enrichment of data with the semantic information can help with communication and mediation between multiple points. The semantic enriched (big) data stored in an open format can be made widely available for the participants of the value chains if it could be further managed by using of cloud computing – the web-based, dynamical IT services. Cloud computing solutions moreover warrant the security on the infrastructure and data level, and also eliminate the need of initial investments in IT infrastructures and shorten the time-to-market.

The usage of open formats may considerably contribute to rising of flexibility and content transfer within supply chains, organized as webs, and simplifying the data transformation into diverse e-business standards.

The drawback of the given approach is the need of its integration into various IT solutions. To be useful it should be supported by the numerous diverse E-business standards as shown in [5].

VI. Conclusion

In the past times the vendors had to exploit earlier period's structured data (stored in SQL OLTP or OLAP systems) in order to analyze customer's attitudes and increase sales. Nowadays, a raising all-embracing connectivity with potentially all stakeholders in supply chains networks results in the possibility of accessing to all needed current data in real-time and also in getting a near-time feedback. This bears the genuine chances for almost immediate improvement of the relationships with the supply chain's stakeholders and therefore increases the agility and ability for just-in-time in reactions to the changing requirements [27]. Accordingly, the high quality decision support becomes possible, which enables achieving optimal performance.

It became feasible to take advantage of this situation facilitated owing to usage of cloud computing and big data by the organizations taking part in the supply chain networks. Nevertheless, the amassed data encountered in the supply chains demand solutions for suited processing of coexistent structured and unstructured data (NoSQL) on the base of appropriate software architectures, and also require a common base as the exchange format of the data shared and exchanged in the supply chains networks.

In the paper we have analyzed the nowadays common solutions for structured and unstructured data storage options for the decision making support. With the opportunity of big data and cloud computing technologies application, the amount of partially unstructured data increases and it needs to be taken into account while making logistic decisions the previous solutions are not enough to cope with the problem dealing with them in the real time or in the near-time.

We see big chances in using such architectures as the layered Lambda architecture (for big data processing), which was designed to cope with the near-time exploitation of big amounts of data arriving. The data exchanged in supply chains has diverse formats, therefore we further propose using an open common data format in supply chains. For an additional advantage a standard solution with Linked Data may be further enriched with semantic information for further support of supply chain collaboration.

It is expected that the application of open Linked Data may substantially support the automated extraction of the information published on the Web by using open standards and additionally describing with semantic meaning and contextual relationships of the data.

We have shown on the 4PL integrator example the need for the integration of social media data for forecasting aims. Also diverse Linked Data from Virtuoso databases can be applied in the decision making support. As a common standard for data exchange between the various IT applications interacting in the logistic value chains we have considered incorporating the semantic concepts associated with the Linked Data into the supply chain management, especially for the aims of the common format integration between the SQL data silos in the value chains using big data.

The usage of the Linked Data by a broker may contribute to the data integration and transfer speed up. The saving of the costs previously needed for the transformations between diverse formats will create the added value for the network participants.

The suggested improvements raise new possibilities for adding value in supply chains. The network effect causes that with increased number of participations the added value for the participants of the network grows.

The Lambda architecture allows merging huge amounts of historical data with near real-time data creating context-oriented data needed for reacting and appropriate responding to different situations like transport events in the logistics.

In the paper we presented the ongoing research work. In the future work the further aspects like economic evaluation of Applying Big Data and Linked Data Concepts in Supply Chain Management should be considered. Also the integration of the further open Linked Data instances from the Virtuoso databases into the supply chains may be investigated.

References

[1] Web Services Activity, W3C Working Group, http://www.w3.org/ws
[2] J. C. Jarillo, "On strategic networks", in *Strategic Management Journal*, 9, 1988.
[3] D. Agrawal, S. Das and A. E. Abbadi, „Big data and cloud computing: current state and future opportunities". *EDBT 2011*, March 22-24, 2011, Uppsala, Sweden. ACM 978-1-4503-0528-0/11/0003.
[4] W3C LinkedData, 2011, www.w3.org/wiki/LinkedData
[5] S. Robak, B. Franczyk, and M. Robak, Applying Linked Data concepts in BPM, *Proceedings of the Federated Conference on Computer Science and Information Systems FedCSIS*. Wrocław 2012. IEEE Conference Publications, ISBN: 978-1-4673-0708-6, pp. 1105-1110.
[6] M. P. Papazoglou, and P. M. A Ribbes, *E-business: organizational and technical foundations, John Wiley and sons*. London 2006, pp.88-90.
[7] M. Kantor, and J. Burrows, *Electronic Data Interchange (EDI)*, National Institute of Standards and Technology, 1996.
[8] E. Dumbill, "*What is big data? An introduction to the big data landscape*", Strata O'Reilly, 11 January 2012, http://strata.oreilly.com/2012/01/what-is-big-data.html
[9] S. Wrobel, „*Big Data – Vorsprung durch Wissen*", Fraunhofer-Institut für Intelligente Analyse- und Informationsverarbeitungsysteme IAIS. Presentation, www.iais.fraunhofer.de, 2012.
[10] I. Mitchell and M. Wilson, "Linked Data. Connecting and exploiting big data", Fujitsu Services Limited, March 2012, www.fujitsu.com.uk.
[11] The Apache Hadoop Project. http://hadoop.apache.org/core/ , 2009.
[12] M. May, "*Living Big Data. Konzeption einer Experimentierplattform*". Fraunhofer-Institut für Intelligente Analyse- und Informationsverar-

beitungssysteme IAIS. Presentation, www.iais.fraunhofer.de, Berlin, 2012.

[13] N. Marz and J. Warren, "*Big data. Principles and best practices of scalable realtime data systems"*. Manning Publications, MEAP Edition, Manning Early Access Program Big Data version 7, 2012.

[14] D. C. Luckham, *Event processing for business: organizing the real-time enterprise*. Hoboken, New Jersey: John Wiley & Sons, Inc., p.3. 2012.

[15] T. Berners-Lee, "On the next Web", Talk on the TED Conference 2009, www.ted.com/talks/tim_berners_lee_on_the_next_web.html

[16] W3C Architecture domain, Naming and Addressing: URIs, URLs, …, http://www.w3.org/Addressing/

[17] W3C Semantic Web, RDF Working Group, *Resource Description Framework (RDF)*, 2004, www.w3.org/RDF

[18] H. Baumgarten, *Das beste der Logistik*. Springer Verlag, Berlin 2008.

[19] S. Chopra, and P. Meindl, "*Supply chain management: strategy planning, and operation*", 3rd ed., Prentice Hall, 2007, pp.427.

[20] A. Matopoulos, and E. -M. Papadopoulou, "The evolution of logistics service providers and the role of Internet-based applications in facilitating global operations" in *Enterprise Networks and Logistics for Agile Manufacturing*, L. Wang, and S.C.L. Koh, Eds., Springer, 2007, pp. 298-304.

[21] D. Allemang, and J. Hendler, *Semantic Web for the working onto-logist modeling in RDF, RDFS and OWL*. Morgan Kaufman, 2008.

[22] OpenStreetMap Project, available at www.openStreetMap.org

[23] DBpedia Project, Free University of Berlin, and University of Leipzig, OpenLink Software, http://wiki.dbpedia.org/About

[24] MastPoint, *CSV-1203, CSV File Format Specification*, Best practice for business-to-business operations, available at http://mastpoint.curzonnassau.com/csv-1203/index.html

[25] Triplify expose semantics!, http://www.triplify.org/Overview

[26] Virtuoso Universal Server, Universal Server Platform for Enterprise Data Integration, Web Services, & Process Orchestration, openLink software, available at http://virtuoso.openlinksw.com/

[27] R. Sethuraman and S. K. Kundharaju, "Top 7 Tips for Big data to optimize Supply Chains. 5 Februar 2013, http://risnews.edgl.com/retail-trends/Top-7-Tips-for-Utilizing-Big-Data-to-Optimize-Supply-Chains86163.