LINKED DATA IN SUPPLY CHAIN MANAGEMENT

Ilona Pawełoszek, Aleksandra Grabińska, Tomasz Turek

Częstochowa University of Technology

Faculty of Managemnt

INTRODUCTION

The concept of logistic supply chain is based on close cooperation among manufacturers, wholesalers, retailers and consumers. This cooperation aims at gaining larger profits for all of the supply chain stakeholders, achieving their market goals, avoiding duplication of efforts and wastage of resources. It is essential for supply chain partners to create a network that is agile and able to respond rapidly to unpredictable changes in demand. Currently, in the era of e-business information technology (IT) is essential in order to compete effectively on the market.

The problem of the appropriate IT solutions for supporting collaboration between supply chain participants is not new and it has been approached with several standards and protocols implemented in numerous enterprise information systems. The applications like ERP (enterprise resource planning), CRM (customer relationship management) or WMS (warehouse management system) etc. contain valuable data that can be utilized by the decision support systems (DSS) , however often there exists an information integration problem. The problems with information interchange are related to issues of data exchange between business partners operating independently designed information systems. From the logistics point of view the key issue is supply chain visibility (SCV), which can be understood as the ability of parts, components or products in transit to be tracked from the manufacturer to their final destination and traced from the customer to the manufacturer. Tracking is the ability to follow the downstream path of a product along the supply chain (Dabbene et al. 2014), and traceability refers to access of product-related records in the upstream stages of supply chain (Bechini et al. 2008).

The goal of providing supply chain visibility is to improve the logistic processes by making data readily available to all of the stakeholders. Supply chain visibility technology promotes near real-time response to market changes by allowing users to take action and reshape demand or redirect supply.

As the scale of collaboration between business partners increases it may require to undertake further steps towards the integration of the IT environment, some possible ways are usage of the loosely coupled e-business solutions based on Web services, cloud computing and the paradigm of Linked Data. The subject of Linked Data application discussed in this paper has very practical dimension, which is reflected in the current focus of GS1 organization on collaboration with W3C [1]. Linked data is more and more often discussed in the context of supply chain management in different industries, for example: in pharma industry for tracking vaccination [2], building distributed networks of supply chain partners on the Web [3], publishing industry and libraries [4].

A review presented in this paper focuses the solutions for providing greater visibility of supply chains by data integration based on the paradigm of Linked Data. By analyzing numerous publications and case studies we summarize best practices for publishing and consuming linked data.

THE NEED FOR DATA INTEGRATION IN SUPPLY CHAIN MANAGEMENT

In the era of globalization supply chains have become very complex and dynamic. The challenges of complexity are related to number of products, facilities needed to perform operations and length of the supply chain. The problem is particularly relevant to global manufacturing companies operating in high-tech sectors of the economy, chemical, pharma and food industry. Many manufacturers face additional challenges related to environmental risks and hazardous materials used for production.

In these industries multiple stakeholders, processes and products generate increasing number of data generating processes. Large volumes of traceability data are recorded at each partner's end with the help of digital technologies. Specific data about deliveries, including destination, size, weight and information about contents, is recorded every day.

The aim of data integration in supply chain is to achieve greater visibility, which is understood as "the ability to know exactly where things are at any point in time, or where they have been, and why."[5]

To make sense of that data, the companies need new strategies, more powerful tools and skills to interpret the numbers, and find the useful insights that are buried in the data.

The main challenges in providing seamless data integration in the supply chain are:

- Large numbers of participants
- Different specificities of participants' businesses
- Heterogeneity of ICT solutions used by participants - enterprises use various data exchange formats and standards,
- Only one up, one down data flow (for example in food supply chain the farmer might communicate with the wholesaler or food processor but not directly with the retailer, a problem even more accentuated in complex supply chains for processed or packaged food)[6].

Supply chains are highly heterogeneous, loosely coupled large-scale networks of entities. Data capture and transmission technologies such as bar coding, Radio Frequency Identification (RFID), Electronic Data Interchange (EDI), wireless networking, and Global Positioning Systems (GPS) produce large amounts of precise logistic data that, if properly managed and shared, can enhance performance and agility of supply chains. To achieve this goal companies require accurate visibility, and a single representation of the supply chain data. This problem cannot be addressed by simple middleware solutions that fail to rectify the fundamental data silo and latency issues[7].

The problems of heterogeneous data formats is partially solved by global standardization organizations and their local member organizations in different countries. One of the most important organizations in this field is GS1 which, during over 40 years of its activity, has defined a detailed Global Product Classification system (GPC), which can be used to identify products that match particular criteria[8]. The example of a standard that facilitates communication in logistic chains to some degree is the Global Trade Item Number (GTIN). There are also other standards that allow to send even more detailed data, one of them is The Global Data Synchronization Network (GDSN) which enables trading partners to globally share trusted product data[9].

Although the standards are effective in unified codification of products, the numbers alone (such as GETIN) are not enough for knowledge transfer from the producer all the way to the final consumer. The issue of providing detailed information on the product provenance and its history (such as when, where and by whom it was packaged, transported) can be very important in case of emergencies (for example contamination of food with bacteria e coli).

Another practical example of a traditional supply chain that exhibits poor information exchange is sustainable wood supply[10]:

One of the wood producers from USA had part of his license revoked but inspite of this fact the firm still applied the FSC[1] label on the wood they were shipping. Some of this wood travelled through the supply chain and unwittingly drifted to UK retailers as certified sustainable product. It is clear that if an integrated information supply network had been available, the chances of this happening would have been decreased, or at least it would have been identified sooner.

Combining product or package identification number with additional information about the origin of the product and the parameters of processes the product has undergone can be useful in many situations. Especially the possibility to access the enhanced data about the product can be valuable for customer to gain knowledge and to ensure about the product's quality.

The IT solutions which provide comprehensive knowledge on the product origin and history can add value to the supply chain and bring numerous advantages for all the participants such as checking reliability of the business partners, support in decision-making processes, providing exhaustive information to the customers. For example it is usually unknown where the raw materials or components of the product came from, this can be important for a customer and vendors for many reasons such as supporting local industry, promoting high-quality goods and services, security of products, etc.

The analyses of the mass amounts of data generated by supply chain processes are valuable source of knowledge and the issue was discussed in many publications describing the application of big data technology which complements with Linked Data [11] and allows for creating even more innovative and intelligent solutions..

However the former problem is how to link the heterogeneous data coming from different participants with the data generated by the supply chain applications so it would be possible to know that the data sources describe the same entity. A systematic and standardized approach to exchange traceability information can be implemented by exploitation of semantic technology – especially in accordance with the Linked Data paradigm. Semantic connections between data sources can be done by exploiting the Linked Data technology.

Linking the data in semantic way can bring new business opportunities and build new sources of competitive advantage for the whole supply chain. The next section discusses practical issues around the implementation of Linked Data.

---

[1] Forest Stewardship Council

THE CONCEPT OF LINKED DATA

Linked Data is one of the possible solutions to implement the concept of Semantic Web in practice. Nowadays, the idea of linking web pages by using hyperlinks is obvious, but this was a groundbreaking concept in the 80'. A similar situation takes place today since many organizations do not understand the idea of publishing data on the web, let alone why data on the web should be linked[12].

The most important advantage and the aim of Linked Data paradigm is to make the data from the Web available and queryable as it is in the case of SQL-based databases. Although appropriate query languages have been developed, the Linked Data approach is not easy to apply due to the large dispersion of data sources and heterogeneity of data schemas.

Linked Data is not a single standard or format but, as Tim Berners-Lee says it is an "expectation of behavior" [13]. Practical application of linked data consists of the following activities: identifying entities embedded in information resources, creating links among them by using named connections, and publishing the relationships as hyperlinks on the web so they can be crawled by machines and people.

Linked Data is generally introduced using its four publishing principles recommended by Tim Berners-Lee[14]:

1. Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

From a programmatic point of view a formal definition can be presented [15 p.211]:

A dataset following the Linked Data principles is a graph G such as $G = (R, L, I)$ in which $R = \{r_1, r_2, \ldots, r_n\}$ is a set of resources — identified by their URI —, $L = \{l_1, l_2, \ldots, l_n\}$ is a set of typed links — identified by their URI — and $I = \{i_1, i_2, \ldots, i_n\}$ is a set of instances of these links between resources, such as $i_i = (l_j, r_a, r_b)$.

As it was stated before, RDF – the Resource Description Framework, is one of the key ingredients of Linked Data, which provides a generic graph-based data model for describing things, including their relationships with other things[16]. Resource Description Framework (RDF) [17] is a simple, yet powerful data model and language for describing Web resources. RDF introduces three basic concepts:

resources, properties and statements. Using RDF, one can make statements about Web resources in the form of triples (subject, predicate, object)[18 p.92]. Using the RDF model, triples can be linked in large graph structures. The nodes of an RDF graph are resources referenced by URIs.

RDF data can be written down in a number of different ways, known as serialisations. Examples of RDF serialisations include RDF/XML, Notation-3 (N3), Turtle, N-Triples, RDFa, and RDF/JSON [16].

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods[20]. Many of the existing linked data sets are available publicly available at no cost. Such data sets are called Linked Open Data (LOD). The most current list of the data sets is published at: http://lod-cloud.net/ and it is usually visualized as a cloud diagram (Figure 1).
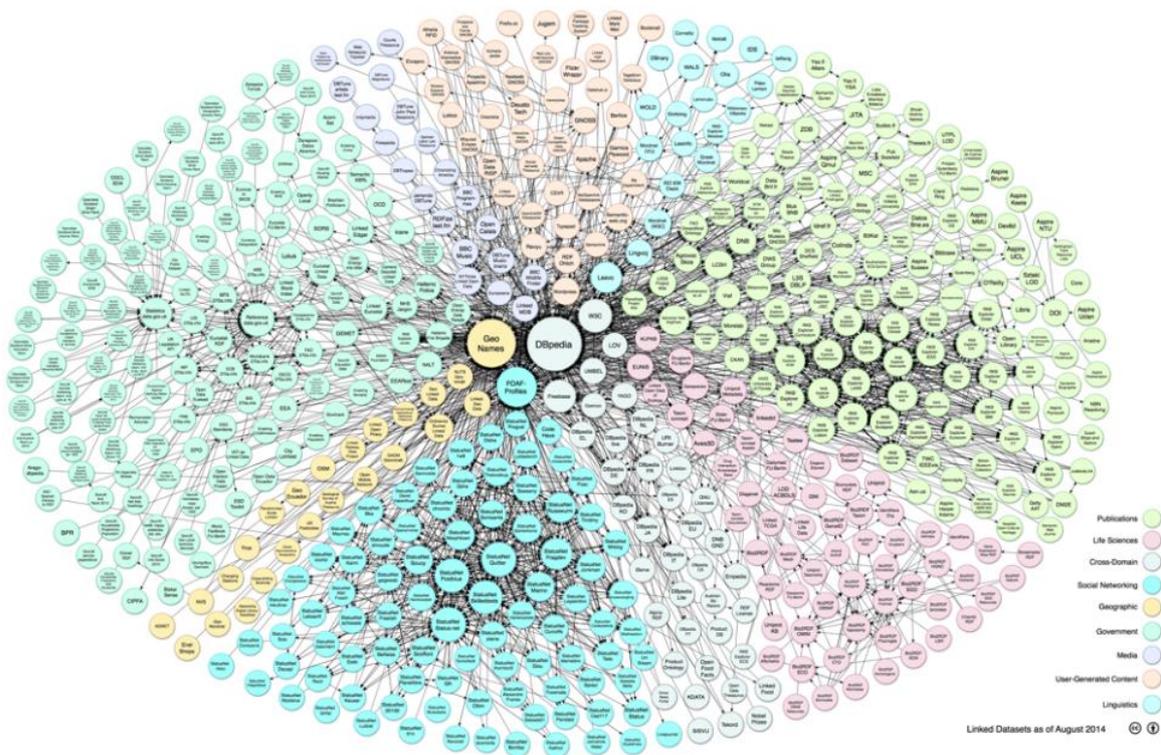

Figure 1. Linked Open Data Cloud

Source: [19]

DATA INTEGRATION WITH LINKED DATA IN SUPPLY CHAIN

In the logistics industry, loads are transported from a start location to a final destination within containers. During the process of transportation, containers are handled in numerous work areas such as ports, warehouses, land, trailers, railway wagons and planes. In case of multimodal international transport the process lasts many days and there is no interface to get detailed information about the transportation. However the customers may want to learn an exact status and position of the transported loads. The problem is partially solved by wireless networks and RFID, but the only information that can be obtained is position and status of the container. With standard hardware solutions there is no way to link the data from RFID tab with other concepts which could be interesting for the customer. For example it is not possible to know the reason of the latency of the delivery, by which factors it was affected. The detailed information can be obtained by making a numerous phone calls to the employees who are responsible for the process [21].

There is a possibility to combine existing standards (such as barcodes and RFID) with Linked Data sources by using online tools which facilitate publishing data on the web in accordance to the rules suggested by Tim Berners Lee. The web applications allow different organizations to export their data in RDF format which is the standard of Linked Data. The data can then be linked to other sources and consumed by many organizations for different purposes. The main goal is to enhance the availability of accurate data about products and services offered by the companies. The availability of this information on the web can bring competitive advantage by offering better support and trusted data for the customer (especially considering the existence of recent legislation such as the European regulation EU1169/2011 regarding food labelling requirements that require detailed and accurate product information to be available online)[8].

In particular, the standardized product IDs published on the web accompanied by Linked Data technologies offer a possibility to connect physical goods (ie. transport units) with digital information on producers, services and consumers.

The four principles of linked data presented in the previous section constitute a base for building good practices in publishing and consuming linked data. The most important detailed rules for the implementation can be summarized as follows [22]:

Proper URIs - HTTP URIs should be used. The next important thing is to use individual namespaces that are characteristic for the publisher. The URIs should not be very complicated, it is a good practice to give them semantic meaning rather than exposing technical details of the imiplementation. An example of properly defined URI is: http://dbpedia.org/resource/Berlin. Whereas too complicated URI, containing unnecessary technical details looks as follows: http://www4.wiwiss.fu-berlin.de:2020/demos/dbpedia/cgibin/resources.php?id=Berlin [23].

Common vocabularies and extending them rather than building new ones. Reusing existing terms is a good practice and there is a set of well-known ontologies developed by the Semantic Web community. Examples of ontologies that can be used are: Friend-of-a-Friend (FOAF), vocabulary for describing people, Dublin Core (DC) defines general metadata attributes, Semantically-Interlinked Online Communities (SIOC), vocabulary for representing online communities, Good Relations Ontology – a vocabulary for describing product, price, and company data, Description of a Project (DOAP), Simple Knowledge Organization System (SKOS) - for representing taxonomies and loosely structured knowledge. If there is no proper vocabulary it is recommended to create it by using commonly accepted terms (for example those defined in ISO or other standards).

The response for the semantic query to linked data sources should be composed in an informative way, containing all the necessary items such as: The description including all triples from the dataset that have the resource's URI as the subject. Backlinks – the representation should also include all triples from the dataset that have the resource's URI as the object. Although such approach may provide redundant data, but it allows browsers and crawlers to analyze links in a bidirectional way. Additional information on related resources can be added, however it should be made carefully to avoid producing excessive RDF file sizes. A needed information can be: author, licensing information and any other metadata which can be valuable from the user's point of view.

Links to other Linked Data sources can be established either manually or by linking algorithms looking for similar content and recognizing its semantic meaning. The first – manual method is more accurate, but also more laborious and therefore is recommended to small data sets. In practice both methods can be used – automatic link generation and then their validation by human.

CONCLUSIONS
Linked Data technologies propose a new solution to dynamic, distributed and complex logistic processes. In this paper, we explore some of the challenges and

opportunities posed by the widespread availability of Linked Data technologies in a supply chain context.

Supply chain visibility which can be improved by Linked Data will also impact other important areas such as: supply chain risk analysis, customer relationship management and partnership between supply chain stakeholders. It should be noted that Linked Data technologies do not replace the legacy solutions but rather have complementary role to the existing supply chain management systems.

However implementation of Linked Data requires developing new procedures by of publishing data and making them an integral part of supply chain operations. Another important aspect of further study, beyond the technical and organizational issues, is providing legal regulations for publishing and using Linked Data. In particular security and privacy issues have to be faced.

REFERENCES

1. http://www.gs1.org/gs1-smartsearch

2. Bhattacharjee P., Solanki M., Bhattacharyya R., Ehrenberg I., Sarma S.E., VacSeen: A Linked Data-Based Information Architecture to Track Vaccines Using Barcode Scan Authentication. [In] Proceedings of the 8th Semantic Web Applications and Tools for Life Sciences International Conference, Cambridge UK, December 7-10, 2015. 2016 http://ceur-ws.org/Vol-1546

3. LUCID – Linked Value Chain Data http://eis.iai.uni-bonn.de/Projects/LUCID.html

4. Antelman K., Wilson K., The Global Open Knowledgebase (GOKb): open linked data supporting electronic resources management and scholarly communication http://doi.org/10.1629/uksg.217

5. http://www.gs1.org/docs/annual_report/GS1_Annual_Report_2012.pdf

6. Brewster C. Wolfert S, Sundmaeker H (2012) Identifying the ICT challenges of the agri-food sector to define the architectural requirements for a future internet core platform. In: Proceedings of the eChallenges e-2012 conference, Lisbon, Portugal, 17-19 Oct

7. http://www.kinaxis.com/en/product/cloud-based-scm-platform/supply-chain-data-integration/

8. Harrison M., Ilic A., Linked Data and the Semantic Web. How are these relevant to users of GS1 standards? http://cocoa.ethz.ch/downloads/2014/06/None_AUTOIDLABS-WP-SWNET-030_01.pdf

9. http://www.gs1.org/gdsn

10. Wallis R., Linked Data a Recipe for Food? March 2, 2012 http://dataliberate.com/2012/03/02/linked-data-a-recipe-for-food/

11. Robak, S., Franczyk, B., & Robak, M. (2013). Applying big data and linked data concepts in supply chains management. Proceedings of the 2013 Federated Conference on Computer Science and Information Systems (FedCSIS) (pp. 1215-1221). IEEE.

12. Bauer, F., & Kaltenbock, M. (2012). Linked open data: The essentials. Retrieved from http://www.semanticweb.at/LOD-TheEssentials.pdf.

13. Coyle, K. (2012). Understanding the Semantic Web: Bibliographic data and metadata. Chicago, IL: ALA TechSource.

14. Tim Berners-Lee. Linked Data. W3C Design Issues, July 2006. http://www.w3.org/DesignIssues/LinkedData.html.

15. Passant A., dbrec: music recommendations using dbpedia. In Proc. of 9th Int. Sem. Web Conf., ISWC'10, pp. 209–224, 2010. p. 211

16. http://linkeddata.org/faq

17. Klyne, G., Carroll, J.J.: Resource description framework (RDF): concepts and abstract syntax. W3C Recommendation. Technical rep. (2004). Available at http://www.w3.org/TR/ rdf-concepts/

18. Fensel, D. Facca. F., Simperl. E., & Toma, I. (2011). Semantic Web Services. Springer p.92

19. http://lod-cloud.net/

20. http://linkeddata.org/

21. Göçebe P., Dikenelli O., Kose U., Bringing Agility into Linked Data Development: An Industrial Use Case in Logistics Domain. Proceedings of the Workshop on Linked Data on the Web, 2015 http://ceur-ws.org/Vol-1409/paper-11.pdf

22. Bizer C., Cyganiak R., Heath T., How to Publish Linked Data on the Web http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/20070727/

23. http://tomheath.com/slides/2009-02-austin-linkeddata-tutorial.pdf