
Big Data Interoperability Challenges for Logistics

PRINCE M. SINGH — MARTEN VAN SINDEREN

*University of Twente
7500AE Enschede
The Netherlands
{p.m.singh, m.j.vansinderen}@utwente.nl*

ABSTRACT. Analysis of real time information can be used to leverage logistic value chains and processes. It enables organizations to become more flexible and efficient. With the continued explosion of available data, challenges are aplenty for logistic companies on how to use such data for meaningful service improvements. Capturing, storing and analyzing such data, extracting useful information, are representative of big data challenges in the logistic value chain. In this paper, we elaborate on the uses of big data analytics for logistics and the related interoperability challenges for logistic companies.

KEYWORDS: Interoperability, big data, logistics, syntactic level, semantic level, pragmatic level

1. Introduction

Logistics industry is moving from a product-related service to an information intensive service, (Sivan *et al.*, 2014; Wegner *et al.*, 2013), where information is a competitive advantage (Wedgwood *et al.*, 2014). Real information about shipment/order locations avoids stock accumulation and makes the supply chain efficient. Abundant data is available these days from various sources, be it real time location of trucks, moving speed and location of deep sea vessels or expected rail disruptions. For strategic decision making the information which logistic companies must consider is rapidly increasing (Wedgwood *et al.*, 2014). Not only there are different sources of data but they are generating data rapidly, which has to be stored and processed in near real time to derive useful information. Owing to its volume, velocity and variety it falls under the domain of big data. According to (Robak *et al.*, 2014), big data is “*data whose scale, distribution, diversity and/or timeliness require the use of new technical architectures and analytics to enable insights that unlock new sources of business value*”. A typical example of big data in logistics would be, processing the board computer information from trucks where each truck sends 2-3 MB’s per minute (Sivan *et al.*, 2014). Usage of big data for enhanced logistic services presents several interoperability challenges for a logistic company. The fact that there are many actors in a logistic value chain add to the interoperability challenges. Although, there are several reports on big data applications in logistics, scarce research literature exists on the big data interoperability challenges (Frehe *et al.*, 2014). The goal of this paper is to discuss the interoperability challenges due to big data usage in logistic. Details over the technologies, methods of extracting, transforming and loading (ETL) of Big data is beyond the scope of this paper. In Section 2 we present background information on big data in logistics. In Section 3 we discuss the logistic process followed by Section 4 where improvement to the process using big data is explained. In Section 5 we discuss the challenges of using big data in logistic. A case study is given in Section 6 to show the state of practice of big data usage in logistics. Section 7 discusses the results and concludes the paper.

2. Background

Successful implementation and benefits of Big Data in logistics has been shown in (Dobrkovic *et al.*, 2015; Sivan *et al.*, 2014; Robak *et al.*, 2014; Ayed *et al.*, 2015). Analytics over Big Data provide useful information for increasing efficiency and in decision making, as shown by (Wegner *et al.*, 2013). Challenges for big data implementation in logistics are discussed by (Mikavica *et al.*, 2014) and (Hofman, 2015). Most existing literature of this topic falls short in two aspects. Firstly, the benefits of big data are not discussed adequately at process level and secondly, interoperability challenges are not discussed at different levels of interoperability. Only in (Janssen *et al.*, 2014) interoperability challenges for big data in logistic are

discussed in detail and a maturity model is given for logistic companies willing to use big data analytics. In this paper, we address the aforementioned research gaps by discussing interoperability challenges at 4 different level for any logistics company. By associating big data analytics with logistic processes will bring the discussion closer to the process level.

2.1. NIST Big Data Taxonomy

With a view to reuse existing research on big data interoperability we refer the work of big data public working group (NIST, U.S Department of Commerce). In (NIST 2015) a detailed big data reference architecture is provided (Figure 1), along with a general taxonomy which is applicable to any big data usage scenario. In the taxonomy there are 7 actors. Figure 1 shows 4 of them, which are most relevant for the scope of this. The *big data provider* collect data from various data providers, cleans and combines the data. It analyses the combined data using software tools and algorithms to create useful insights, then provides access to the data consumer to make use of these insights. For ease of access and usage these insights are presented in the form of charts, graphs, heat maps or any other visualizations desired by the data consumer. To perform these activities (collection, preparation, analytics, visualization and access) the *big data application provider* can use the infrastructure provided by the *big data framework provider* (which is essentially an IaaS provider). A *data consumer* can have more than one big data application provider which in turn can have multiple data providers and big data framework providers. A LSP (logistic service provider) is a company which receives shipment orders from customers and uses its resources or those of its partners (or a combination) to fulfill orders. It is assumed that every LSP wants access to *useful information* for better efficiency. For the scope of this paper, we use the term LSP to include the concepts of 4PL and 3PL (Singh *et al.*, 2015a).

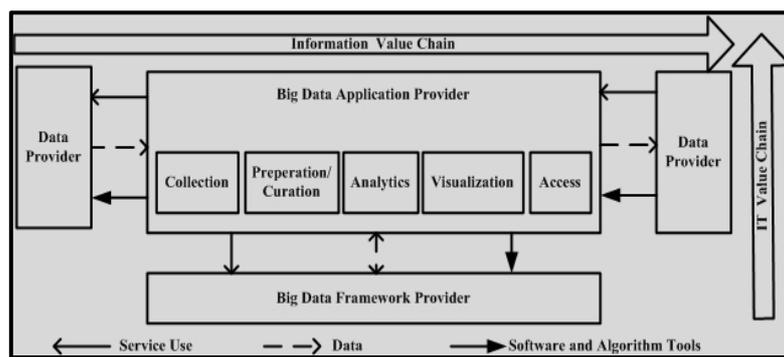


Figure 1. Adapted version of NIST Big data reference architecture (NIST 2015)

2.2. Existing Big Data Applications in Logistics

Contextual data from various sources can be used in demand prediction, optimizing transport cost and intelligent warehousing. A good starting point to study existing application of Big Data in logistics is (Frehe *et al.*, 2014). Frehe *et al.* review big data application in logistics and supply chains based on extensive literature search and interviews with LSPs. Table 1 enumerates the most relevant big data implementation in logistic. In (Wedgwook *et al.*, 2014) three main areas are indicated where big data analysis can be used for positive business outcomes. These areas are, (a) Customer analytics and loyalty marketing (b) Capacity and pricing optimization and (c) Predictive maintenance analysis. In (Wegner *et al.*, 2013) the possible uses of Big Data for logistic companies are discussed in more detail. A selection of the main ones are as follows - (a) Last mile optimization (b) Resource capacity planning (strategic i.e. long term and operational i.e. daily) (c) Customer loyalty management and Service improvement. (d) Supply chain risk management. (e) Strategic planning (f) Operational capacity planning. (g) Risk evaluation and resilience planning. For strategic decisions a high data volume from a variety of sources is to be combined in order to support investment and contrasting decisions. Whereas, operational decision making requires continuous flow of real time data (Wedgwook *et al.*, 2014).

Ref.	Data Source	Used for	Related process
(Wedgwook <i>et al.</i>)	orders, traffic data	planning, resource allocation	planning, order consolidation
(DHL, 2016)	social media, blogs, weather etc.	risk evaluation, resilience	disruption handling
(Sivan <i>et al.</i> , 2014)	board computers	driver behavior monitoring, idling of trucks	monitoring, order review
(Ayed <i>et al.</i> , 2015)	gps, rfid scanner	monitoring and tracking of cargo	monitoring
(Frehe <i>et al.</i> , 2014)	sensors, gps	optimize refueling	monitoring
(Frehe <i>et al.</i> , 2014)	websites	better price negotiations	pre-planning
(Frehe <i>et al.</i> , 2014)	telematics, drivers work patterns	reduce risk of accidents	planning, monitoring, order review
(Frehe <i>et al.</i> , 2014)	board computer	advise to driver, maintenance schedule	monitoring
(Frehe <i>et al.</i> , 2014)	GPS, telematics, truck data	minimize waiting time, resource utilization	monitoring, disruption handling

Table 1. Existing big data implementations in logistics

As an example of operational optimization, historic traffic data can be combined with weather forecasts to create an optimum route/best route at any time. As in (Dobrkovic *et al.*, 2015), real time AIS (Automatic Identification System) data is collected and analyzed to predict the arrival of deep sea vessels at the harbor. Based on this insight, LSPs can better plan when to send trucks to the harbor.

3. The Logistic Process

An LSP performs numerous sub-processes, activities and interactions with other actors in the domain. Based on interviews with LSPs we gained further insight into the 4 sequential basic processes. Figure 2 shows a simple activity diagram of the main logistic processes (Singh *et al.*, 2015b). This elaboration is essential to discuss the usage of big data at the process level, otherwise we would be referring to processes which we have not described in the text.

Below we discuss 4 of them, which can benefit most from big data analytics, in detail. (1) *Planning* is preceded by preplanning, which is open to modifications later. The pre-plan is used to gauge the price of shipment and send invoice to the customer. An LSP can use historic data, order characteristic and future demand expectation etc. to decide the most optimal price of the shipment. At a pre fixed time the LSP plans the best route for the shipment based on contextual information like, future demand, current fleet, expected disruption etc. Usually the pre-planning and planning is done manually in small companies. An LSP frequently combines different orders referred to as Order consolidation to achieve an optimal loading strategy. (2) *Disruption Handling*. Events like rail works, extreme weather condition, accidents can adversely affect ETA of shipments. In these case a new plan might be required or the current plan has to be revised. This is called disruption handling. Based on historic data, an LSP can predict how long a certain disruption will last. Using historic data an LSP can (a) Pre-empt a disruption (e.g. break down of truck), (b) eradicate the adverse effect of the disruption by devising a new plan. DHL Resilience 360, is a risk management solution from DHL (DHL). All disruptions along the route, their cause and location are displayed on the map in near real time. Using this information planners can re-plan critical and future orders. (3) *Order monitoring*. The location of orders and carrier (truck, barge, ships etc.) is monitored by the LSP as customers are interested in knowing the exact location of their shipment at all times. Moreover, data collected during shipment helps determine service and operational benchmarking. (4) *Order Review*. After the order is delivered to its final destination the LSP reviews the orders. Performance of the carriers, total CO₂ emission, and departure from actual plan are stored. It is a part of management reporting which is a priority area for improvement in logistics. In Table 1, existing big data applications in logistics are mapped to the processes in Figure 2.

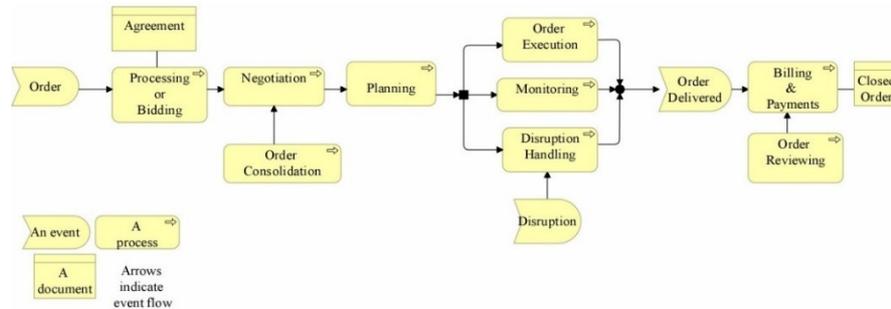


Figure 2. Logistic processes of an LSP

4. Levels in Big Data Interoperability for Logistics

There exist many definitions for interoperability. For the scope of this paper we use IEEE definition, *interoperability is the ability of two or more different systems or components (i.e. software components, processes, systems, business units, etc.) to exchange information and to use the information that's been exchanged*. With respect to interoperability the *variety* of data is the most important aspect (Singh *et al.*, 2015a). We discuss Big Data interoperability at 4 levels, namely technical (or data), Syntactic, semantic and pragmatic, as indicated by (Janssen *et al.*, 2014). Without achieving interoperability at each of these levels it would be difficult for an LSP to achieve the promised benefits from big data.

Technical Level. According to (Janssen *et al.*, 2014) LSPs aiming for big data interoperability have to start at technical level. They first need the hardware and software to read and store diverse data streams. They would need new database implementations suitable for big data. To provide insights these storages would have to work together with the existing traditional database. For ad hoc queries, the algorithm would have to be changed so that data from different (and now distributed) files systems is retrieved and analyzed. Moving to such an extended databases would not be easy for big LSPs have a matured and robust IT infrastructure in place. Other challenges at the technical level such as coding and decoding messages from different data sources can be met by following the existing standards.

Syntactic Level. The classical data *variety* challenge of big data (Mikavica *et al.*, 2014) is quite explicit in logistics. Data stream are of various types their content and format is diverse, i.e. structured (e.g. GPS), semi structured (e.g. weather data), quasi-structured (e.g. websites, AIS data) and unstructured (e.g. tweets) (Robak *et al.*, 2014). Big retail chain in Netherlands forecast future demands and plan warehousing and truck movements accordingly. The retail chains also have to

incorporate data from partner carriers so see their resource allocation patterns and availability. Such a cooperation is not possible if data is not shared using a common format. For a LSP working on different modalities, there are additional problems of syntactic interoperability. It has to decode data from different sources and for different mode of transport too. Consider a shipment corridor that consists of road as well as inland shipping. To determine congestions on this route at any given time, data needs to be combined from sources providing traffic information for roads with inland terminal availability to gauge how fast/slow this route is. For this (Hofman, 2015) proposes a data collection engine, which is tasked with the work of collection, mixing and checking the trust worthiness of the data. To ensure that the indicators are consistent, it is required that the LSP and the data collection sources have used a common, agreed standard with clear rules. Website of infrastructure providers (like port authorities, airports, terminals and rail network owners) are a source of big data. This data can be used to recognize patterns for availability, congestions and disruptions along routes and corridors, as demonstrated by (Dobrkovic *et al.*, 2015) using AIS data. These websites have different layouts and formats and custom mapping or plugins are required for each website.

Semantic Level. As an example of data mismatch at semantic level, let's say a warehouse sensor records the time when a packet leaves the warehouse and refers to this time as 'shipment departed'. A different warehouse refers the time a packet was loaded in the truck as 'shipment departed'. In the absence of a standard data collection policy, if the data from the two warehouses are combined it will result in anomalous conclusions. Semantic interoperability implies that information is interpreted in a consistent manner by different users. Deduction based on one dataset in a certain context might not apply in a different context, as found in the study of illegal rooming houses in New York (Baclawski, 2014). Therefore, metadata model, consistent through all the processes is needed facilitate semantic interoperability. This would enable consistent interpretation of the information, identify trends, and unearthing new insights. Deduction rules should be consistent. For example, bad weather forecast implies delay in shipment due to road congestion but not vice-versa. It frequently happens that data from different sources indicate opposite future trend or situation. It will require machine learning techniques to deal with such situations. If high congestion on road is reported by carrier website and also by social media, how can it be determined that they both refer to the same cause of disruption? Most of these challenges can be addressed by using a data portfolio (Janssen *et al.*, 2014).

Pragmatic Level. Most of the planning, negotiations and decision making in logistics is done manually. Integration of insights from Big Data with existing human knowledge, expertise and intuition across the organization demands pragmatic interoperability. Consistent interpretation and decision making across department and business units has to be achieved. Other challenges at the pragmatic level include, integration with existing BI applications, finding appropriate staff with

required knowledge and analytical skills & organization of training courses. In the case of future disruption along a route, a truck driver has to be alerted and also the route planner. They have access to different systems with different interfaces. How much and at what time they must be informed about the disruption? Business rule, data access policy and privacy policy are some other challenges.

5. Solution Approach

Notwithstanding the promise of big data benefits for LSPs their core business is logistic and surely not data science or IT. It might be one of the main reason why (Frehe *et al.*, 2014) did not receive much enthusiasm from German logistic companies over big data implementations. Big data infrastructure require new investment and advanced IT skills among employees. For some LSPs the profits in return might be much less or take longer as compared to the investments needed for setting up the infrastructure and its maintenance. As a solution LSPs can use Big Data Application (BDA) providers (Figure 1) to provide big data analytics. This implies that the LSP would primarily be concerned about semantic and pragmatic interoperability. If the BDA provider imposes a data standard, then the LSP has to make a mapping onto its internal data model. Since different BDA providers may use different standards, there may be multiple mappings. LSPs can also use the IaaS platforms provided by Big Data Framework Providers (BDFP) (Figure 1) as IT infrastructure for big data thereby reducing the installation and maintenance cost. In such situations, the LSP is only the data consumer (Figure 1) having a set of requirement over the kinds of analytics/information is needed (e.g. road congestion, historical demand, future disruptions). An LSP can have more than one BDA provider for each mode of transport. Nevertheless, the company would still have to make combine all this data to make a common derivation based on these analytics for an intermodal route.

In Figure 3 we show a representative diagram showing the consumption of big data analytics by LSPs. It's highly unlikely that each data provider would be willing to switch to a common standard for data access owing to business demands. Another possible approach which can be adopted by an LSP is to develop its own infrastructure for big data analytics. It can be an option for big logistic companies like DHL and shippers like Procter & Gamble. Such companies own shipment infrastructure, warehouse, have high demand and bargaining power with other logistics infrastructure providers, like terminal operators. A step by step approach toward big data interoperability can be followed by these companies as indicated by (Janssen *et al.*, 2014).

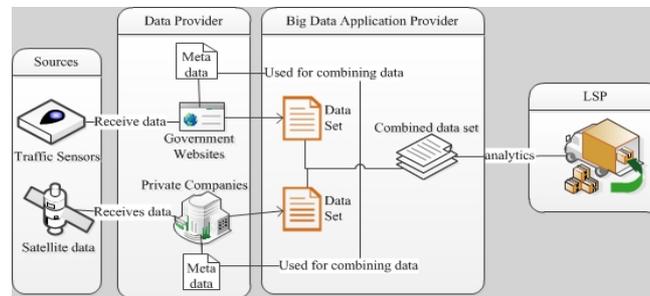


Figure 3. Big Data reference architecture for logistics

6. Case Study

In this section, we present a case study over Simacan, a big data application provider based in Amersfoort, NL. The purpose of this case study is to show how the interoperability challenges are addressed by Simacan. Simacan is an important partner in two major projects. (1) *A58 Project*. The goal of this project to detect and analyze traffic jams on a Dutch highways. Simacan's task in this project is collection, preparation and curation of traffic data. The influx of data is in the order of magnitude of 1 million records per minute. Analytics on the data is done by other companies. These companies then advice truck drivers and cars for optimal driving speeds. Data is collected via the following sources, (a) from TomTom (b) from Flitsmeister (www.flitsmeister.nl) (c) Rijkswaterstaat, i.e. Dutch government (d) NDW, government operated brokerage service for traffic data (semi-public data). The method of obtaining each of this data is different while for (a) it involves receiving a continuous data stream, it is downloading a text file a server in (c). The type of location referencing from these sources are different and each of them has to be mapped to OpenLR, which is used as a standard by Simacan. In case a new data source is added a new mapping would be required. (2) *Simacan Control Tower* is a product of Simacan used by several companies in the Netherlands. One of the use cases of the Simacan Control Tower is the distribution of goods to stores of a large retail chain in the Netherlands. Goal of the Simacan Control Tower is to monitor and control the distribution operation of the retailer. In this process, they provide insight into ETA's for the stores as well. Planned routes and stops of trucks are known well before departure on the day of delivery. For the ETA's provided by Simacan, they subscribe to live traffic data feed from TomTom and also provides access/visualization for the planners at the retail chain. They can monitor the ETA of trucks at stores in real time and plan their operations. Currently, Simacan does not provide an integrated planning module, which can suggest the optimal plan and routes for the truck. However, real-time planning rescheduling and pre-trip planning tools are features that are possibly added to the product in the future.

7. Conclusion

In the highly dynamic world of logistics, LSPs have to develop *dynamic capacities*. Such capabilities are required to identify useful data sources and utilize them for competitive advantage and new value creation (Janssen *et al.*, 2014). Big Data is an area of untapped potential for improving resource utilization, reducing and anticipating risk and improving customer experience (Hofman, 2015). LSPs have to recognize areas of process improvements and gain access to required information for better planning as shown by us in Sec 3. But first, they have to ask the right questions. What are the goals which require a big data implementation? What process have to be improved? How will the success of the implementation be measured? Is there an existing policy and governance mechanism? It might occur that *some* of insights and improvements can be achieved even without big data analytics, for e.g. from historical data. For better semantic interoperability of big data, LSPs can make use of a data portfolio (Janssen *et al.*, 2014). A data portfolio shows all the data sets, along with their intended use, ownership, quality and business value. A data portfolio also enables combining different sets to obtain advanced analytics. It is open to debate how much resources, small to medium sized LSPs are willing to invest in a full fledged big data policy and infrastructure. Furthermore, actual monetary benefits from Big Data analytics for small to medium LSPs are still to be researched. But the presence of big data application providers and big data framework providers can reduce the investments on infrastructure considerably.

In this paper we have highlighted the use of big data analytics for logistic. Along with the potential benefits, there are interoperability challenges which we discussed at four levels. Using a case study we highlighted the challenges faced by a big data application provider in combining diverse data sources. We conclude that a step by step approach needs to be followed by LSPs to develop their capabilities for big data usage and analysis. Not before, they have a clear idea of why they are pursuing such a strategy and to what ends. This paper opens several research areas for future research. It has discussed big data from a process viewpoint rather than a technical one, which has been indicated as a research focus in previous papers (Frehe *et al.*, 2014).

8. References

- Ayed, B.A., Halima, B.M., Alimi, M.A., "Big data analytics for logistics and transportation", *ICALT 2015 Conference*, IEEE: 311–16, 2015.
- Baclawski, K. "Semantic interoperability for big data", Presentation, 2014. <http://tinyurl.com/jgksu48>
- DHL Resilience 360, <http://tinyurl.com/hysutmu>, accessed on 20-02-2016.

- Dobrkovic, A., Iacob, M.E., Hillegersberg, J., “Using machine learning for unsupervised maritime waypoint discovery from streaming AIS data”, *i-Know 2015 Conference*, Graz, Austria, ACM, 2015.
- Frehe, V., Kleinschmidt, T., Teuteberg, F., “Big data in logistics – Identifying potential through literature, case study and expert interview analyses”, *Lecture Notes in Informatics*, Springer, 173–86, 2014.
- Hofman, W., “Data collection architecture for big Data – A framework for research agenda”, *BDEI@IWEI 2016 Workshop*, Nimes, France, CEUR-ws.org vol. 1414, 2015.
- Janssen, M., Estevez, E., Janowski, T., “Interoperability in Big Open and Linked data-organizational maturity, capabilities and Data portfolios”, *IEEE Computer* 70(10): 44–9, 2014.
- Mikavica, B., Kostic-Ljubisavljevic, A., Dogatovic, V.R., “Big Data: Challenges and opportunities in logistics systems”, *2nd Logistics Intl. Conference*, Belgrade, 185–90, 2014.
- NIST, “Big data Interoperability Framework: Volume 6. Reference Architecture”. <http://dx.doi.org/10.6028/NIST.SP.1500-6>
- Robak, S., Franczyk, B., Robak, M., “Research problems associated with big data utilization in logistics and supply chain design and management”, *Computer Science and Information Systems* 3: 245–9, 2014.
- Singh, P.M., van Sinderen, M.J., “Interoperability challenges for context aware logistics services – The case of synchromodal logistics”, *BDEI@IWEI 2016 Workshop*, Nimes, France, CEUR-ws.org vol. 1414, 2015a.
- Singh, P.M., van Sinderen, M.J., “A common data model for synchromodal logistic planning”, Technical Report, University of Twente, 2015b.
- Sivan, P.A., Johns, J., Venugopal, J., “Big data intelligence in logistics based on Hadoop and MapReduce”, *IJRSET* 3(3): 2634–40, 2014.
- Wedgwood, K., Howard, R., “Big Data and analytics in travel and transportation”, IBM Big Data and Analytics White paper, 2014.
- Wegner, M., Kuckelhuis, M., “Big data in logistics”, DHL customer solutions and innovations, Report, 2013.