Proceedings of 7th Transport Research Arena TRA 2018, April 16-19, 2018, Vienna, Austria

Big data analytics in e-commerce logistics: Findings from a systematic review and a case study

Eleni Zampou [a], Christina Milioti [a], Aggelos Liapis [a], Vega Rodrigalvarez [b], Florian Flocke [c], George Dimitrakopoulos [d], George Bravos[e]

[a]*ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS, Athens, Greece*
[b] *ITAINNOVA, Madrid, Spain*
[c] *FRAUNHOFER IML, Dortmund, Germany*
[d] *INTRASOFT INTL, Brussels, Belgiium*
[e] *ITML, Athens, Greece*

Abstract

Digital evolution has significantly changed consumer shopping habits and expectations resulting in a major growth of e-commerce. The immediate outcome of this growth was the creation of a dynamic and turbulent environment with increasing density of the distribution network. This environment consists of many delivery points, multiple delivery channels and last-mile delivery requirements. Due to its complexity and the proliferation of data challenges, e-commerce logistics is an area where the application of Big data analytics can be proven to be extremely fertile. Despite the growing interest, there are limited studies that investigate the role of Big data. By using a design science approach, we clarified the current e-commerce logistics practices as well as the envisioned ones that can, to a large extent, be supported by appropriate big data technologies. We concluded to a set of business requirements that express the needs towards the e-commerce logistics. Then, the requirements were translated into a set of use case scenarios to demonstrate how they could be supported by big data analytics. We conclude by proposing a conceptual architecture of a big data analytics artefact that could cover the e-commerce logistics requirements.

Keywords: big data, business analytics, e-commerce logistics, collaboration, requirements

1. Introduction

Digital evolution has changed consumer shopping habits and expectations with the number of internet users that become internet shoppers to increase steadily on a yearly basis. In the last decade, the e-commerce sector is booming as almost all growth in retail sources from e-commerce and about 57% of European Internet users shop online (E-commerce Europe, 2016)). For example, the European B2C e-commerce sales have been growing steadily since 2011 resulting in a European B2C e-commerce turnover of €509.9bn (E-commerce Europe, 2016). Customers today demand a high level of service from retailers, regardless of how they buy products or receive delivery. Consumers expect shopping to be convenient and expectations are still rising: 40% are looking for even easier shopping across on- and offline channels, 41% expect improved customer service, 45% want improved delivery service, and 46% ask for easier return and refund (E-commerce Europe, 2016). Therefore, the retailers have to deliver a seamless customer experience at every touchpoint, maximize sales across every channel and device, and live up to their promises regarding product availability and delivery.

Meanwhile, the omni-channel growth has created new challenges for the retailers such as inventory visibility, multiple channels integration, speed of delivery and ease of returns. Whether their business is online, in physical stores or both, all retailers face similar challenges. Looking at the UK and Germany, we see that more than 65% of shoppers aged 21 or younger prefer to have both a physical and online experience (E-commerce Europe, 2016). The increase in the variety of channel formats, and the progression from single, to multi, and then to omni-channel formats has made shopping and buying more convenient for consumers, but trickier to manage for logisticians—both upstream suppliers and downstream retailers. Moreover, cooperation schemes where traditional brick and mortar retailers merge their powers and their networks with online retailers have started to emerge. Just recently, Amazon has bought Wholefoods and it will extend its business to the physical world (Aziza, 2017). This new trend will disrupt the current way that the market works and will also demand more effective and efficient logistics. Retail formats that deliver on this complexity will therefore be more successful in the future (E-commerce Europe, 2016).

Responding to the emerging trend triggered by e-commerce growth, the transportation industry is undergoing a major transformation nowadays. The e-commerce initially and the omni-channel growth lately has created a dynamic and turbulent environment, where the density of the distribution network, comprising many delivery points, multiple delivery channels and last-mile delivery requirements, significantly increases the distribution cost (Boyer, 2009; Hübner et al., 2016). Stores are becoming fulfilment centres, serving as pick-up locations for online orders and customers require more dynamic delivery services for the last mile, rendering this process a challenging task (Jeseke, 2013). Home delivery services, which are usually preferred by the online consumers, contribute to the atomization of parcel flows, thus causing problems within the urban areas. In this context, complexity has also raised by the variety of distribution requests, e.g. orders and returns from various consumers at different points. Therefore, e-commerce logistics turned to be a key area of innovation and one of the European Commission's key policy areas (E-commerce Europe, 2016).

Taking into account the growing interest for e-commerce logistics, previous studies have recognized this area as one where big data can be extremely valuable. Previous contributions suggest the need for identifying ways of how big data can assist companies to reap the envisioned benefits (Chen et al., 2012, Fosso Wamba et al., 2015, Sharma et al., 2014). Although there are some existing industrial business use cases, implementation of big data techniques in supply chain management and transport (Castiñeira and Metzger, 2018) and especially in e-commerce logistics, appears to be an untapped asset across industries (Addo-Tenkorang and Helo, 2016). Due to the continuously changing e-commerce environment and the limited application of big data solutions at the e-commerce logistics, further research is needed regarding the actual design and implementation of Big data analytics in a real-world context and identifying their impact on e-commerce logistics.

Building up on the aforementioned evidences, this study aims to shed light on the emerging landscape of e-commerce logistics and to investigate the role of big data analytics in transforming the current e-commerce logistics practices and bring value on all the involved stakeholders 3PLs, online retailers and final consumers. More specifically, following a design science research approach, that employs systematic literature review, individual in-depth interviews with selected industry experts, on-site observation and discussions with a 3PL user, we have clarified, in sufficient detail, the current practices and the envisioned ones that can, to a large extent, be supported by the big data technologies and finally, we concluded to a set of requirements that express the needs towards the e-commerce logistics from the side of the (end-)users. The requirements are then translated into a set of use case scenarios in order to show how these requirements could be covered by a big data analytics artefact. We conclude by proposing a conceptual architecture of a big data analytics artefact that could cover the e-commerce logistics requirements.

2. Research background

As e-commerce expands at fast rates of growth (Ecommerce Europe, 2016), new business processes are being developed and also new consumer patterns of behaviour are being built. Due to the challenges and opportunities created by the information revolution, there has been an increasing emphasis on big data analytics (BDA) in e-commerce in recent years. In the e-commerce context big data analytics (BDA) can be used to improve decision making in all activities involving infrastructure and operation on one hand, and consumers' behaviour and satisfaction on the other (Miller, 2013), thus archiving a better matching between supply and demand. Big data entails the capturing of big volumes of data from different sources and the processing and analysis of them in a real-time environment. A review of big data analysis in e-commerce can be found in Akter et al. (2016). Personalized services, dynamic pricing, predictive analytics, supply chain optimization and visibility are some of the core big data application areas in the e-commerce field (Akter et al., 2016; Koutsabasis et al., 2008; Mehra, 2013). It has been observed that companies that incorporate BDA in their business show a 5-6% increase in their productivity (McAfee and Brynjolfsson, 2012).

Many studies show that customers consider the logistics performance as an important factor of e-commerce (Agatz et al. 2008; Bask et al., 2012; Esper et al. 2003). Distribution of goods to customers and in particular, last mile operation is probably the most demanding process in e-commerce (Moroz and Polkowski, 2016). This is because in e-commerce customer distribution deals with frequent personalized orders resulting in high costs and difficult to manage processes. Outbound delivery costs are high because customer orders do not fill a truck resulting in low utilization of delivery trucks and personnel and longer routes. In addition, re-distribution, or returns system, is considered to be one of the most problematic and costly activities in e-commerce logistics (Hjort and Lantz, 2016). An efficiently organized return process, as part of the distribution system, can help to retain the customer by minimizing his inconvenience and reducing pick up and return handling costs.

There is a growing body of the literature that discusses specific aspects of e-commerce logistics e.g. new distribution practices (e.g. pick-up points, automated lockers, crowd-sourcing), return policies and their impact under various contexts. Morganti et al. (2014) compare the alternatives to home delivery that have been developed by French and German parcel delivery operators which developed pick-up points in stores and automated lockers networks. Moroz and Polkowski (2016) extend the analysis for pick-up points and parcel machines by taking into account customers' environmental attitudes. An effective large-scale mobile crowd-tasking model, in which a large pool of citizen workers is used to perform the last-mile delivery, was suggested as an alternative for last mile delivery (Wang et al., 2016). Hjort and Lantz (2016) first attempt to empirically analyse and describe the effects of returns policies on consumer behaviour and the moderating effects of the policies on profitability. In summary, this stream of research investigates practices that will enable e-commerce retailers to enhance their distribution processes and decrease the delivery cost.

The above applications and trends in e-commerce logistics can be efficiently supported by big data analytics (Swaminathan, 2012). Big data collection and analysis can provide effective decision support in three main areas of logistic applications: operational efficiency and network planning, customer experience and new business models (Jeseke et al., 2013). Previous studies have highlighted the role of Big data analytics in the supply chain (Addo-Tenkorang and Helo, 2016), and stressed the importance for companies to acknowledge it as a strategic asset to be understood and integrated holistically (Wang, et al., 2016). Specific use cases that exploit the value of big data analytics regarding distribution process and contribute to operational efficiency have been identified in the literature. Real time route optimization (Pillac et al., 2013; Fabian and Christian, 2012), demand forecasting, crowd based pick-up and delivery (Jeseke et al., 2013), dynamic inventory routing (Sage, 2013; Wang et al., 2016) are some of the main big data applications in the logistics domain. Based on the literature review findings, there is room for improving logistic processes and especially e-commerce logistics using BDA.

3. Research approach

Following up on our research objectives and the research questions described above, we adopt as methodological backbone the design science paradigm (Jones and Gregor, 2007; Hevner et al., 2004; Peffers, 2007). The particular approach seeks to solve problems through innovative and creative development of artifacts that apply, and test, theories and methods through an iterative process of development (Berente and Lyytinen, 2006).

Initially, we defined the specific research problem and justified the value of Big Data Analytics in the area of e-commerce logistics. The identification of the research gaps was based on literature review combined with semi-structured interviews with industry experts regarding business practices. A series of in-depth interviews were conducted from 20/2/2017 to 10/3/2017 in five companies to get insights about industry needs. Interviewees were selected based on two criteria: the nature of their business (pure online vs click and mortal companies) and the cooperation or not with a 3PL company. In order to conduct the interviews an "interview protocol" was developed. Via the protocol the follow areas were examined: a) Demographic characteristics of the companies, b) Overview of the current situation in e-commerce logistics, c) Future scenarios in e-commerce logistics and the role of big data technologies and d) The attitude towards collaborative logistics practices in e-commerce area. The interviewees were three retailers in the sector of electronics (one of them pure online and two of them click and mortal) with presence in all country, as well as a courier company which is the leading courier company in Greece and a 3PL company that focuses on e-commerce logistics. Through this method the number of problematic areas in the current practices, accompanied by some drivers and barriers towards the adoption of collaborative logistics practices, and many challenges that should be taken into consideration for both the design and the implementation of big data analytics artefacts were identified. The results of this step are reported in Section 4.1.

By analysing and consolidating the insights retrieved towards the process described above, we concluded into a set of high level business requirements (discussed at Section 4.2). In order to further elaborate the business requirements and proceed with the design of the big data analytics artefact, we identified a set of use case scenarios that cover the aforementioned business requirements (Section 4.3). The scenarios were defined by taking into account the industrial context of our research (brick and mortar retailers, online retailers as well as manufactures/suppliers). In order to set the specific objectives of the big data analytics artefact and move to the design and validation of the proposed solution, we worked in close collaboration with two representative retailers, a brick and mortar retailer and an online retailer and with a 3PL company. More specifically, the brick and mortar retailer is a major Greek retail chain and its supply chain consists of three central warehouses and a total of 240 stores (190 supermarkets and 50 CashandCarry). The online retailer is one of the first and major retailers in Greece and the 3PL is one of the biggest Greek retailer in the e-commerce logistics. In this way, we defined the objectives of a big data analytics artefact from the problem definition and knowledge of what is possible and feasible.

The next step involved the design and development of the artefact. The artefact in our study is a Big data analytics infrastructure. The design and development process follows that of an Information System development project. It started with a requirements-gathering process, in which a diverse set of potential users participated (as presented before), resulting in requirements documentation, which was later used for designing a conceptual architecture (presented at Section 5).

4. Business requirements and use case scenarios in the e-commerce logistics

4.1. In-depth interviews and literature review insights

In this section, we present the insights collected towards the in-depth interviews and discuss their relevance to the findings that came towards the literature review. It is important to note that the evidence collected during the interviews and the literature findings converged since the companies selected represent not only local but also multinational or extroverted companies. Therefore, we could claim that the business requirements presented in the next section are generic enough despite the country specific challenges. First, interviewees stressed the fact that big data technologies can collect data from various sources (e.g. social media data, sales data, locations data and geolocation data) and provide solutions that can decrease the delivery cost and also improve the efficiency of the distribution process. Exploitation of big data in logistics requires matching data sources, analytic procedures and business understanding. According to the literature, there is a huge potential for improving operational efficiency by using big data analytics. Operational efficiency includes both last mile optimization, as well as network planning. Last mile efficiency can be considerably improved by real time optimization and consolidated pick-up and delivery. A second area of big data applications in logistics regarding operational efficiency field is network planning, which includes decisions concerning warehouses, distribution centres and custom-built vehicles (Jeseke et al., 2013).

During the interviews, the main characteristics and the problematic areas of e-commerce and logistics market were discussed. Almost from the very first discussions it was evident that the competition in the market of e-commerce is harsh, which constitutes one of the most significant barriers for the implementation of collaborative logistics practices and the data sharing among supply chain partners. The geographic diversity of the country and its small population density, the problematic zip code system, the national orientation of most of the retailers' companies, some special characteristics of the most 3PL companies and couriers and a number of legislative restrictions and the effects of the on-going fiscal and economic crisis are the main problematic areas that have been identified. The aforementioned aspects result in a high distribution cost that is even higher at the last mile part of the delivery process.

Last mile delivery was recognized as the most expensive part of the delivery process and as the most challenging one as it directly affects the consumers' perceptions about a brand by all interviewees. In the Greek market, this cost is also increased by the phenomenon of "cash on delivery", that is a payment method preferred by almost 75% of online consumers. By selecting this payment method, the consumer keeps the right of not to accepting the delivery that has a result an increased number of undelivered orders that needs to be returned. Another aspect that needs to be taken into account for the high last mile distribution costs is the oligopoly market of couriers' companies in the Greek market. One of the key players holds 40% of the market share and the other five couriers share the rest. However, couriers play a significant role in the e-commerce logistics chain as they are responsible for the last mile distribution and the respective cost. Moreover, they can provide traceability data of their orders that can support a series of advanced analytics in e-commerce logistics and can also release a series of dynamic delivery services that can be used by consumers.

Last mile distribution can be considerably improved if we use real time information from different sources (traffic data / sensors / real time events) to dynamically optimize routing and provide drivers with directions on the spot (Pillac et al., 2013; Fabian and Christian, 2012). In e-commerce deliveries, dynamic routing is even more challenging due to the fact that orders change unpredictably and dynamically (Du et al., 2005). Dynamic vehicle routing allows for taking into account both the delivery and the new return requests. As indicated by three out of five interviewees, re- distribution or returns system is considered to be one of the most problematic and costly activities in e-commerce logistics and thus, dynamically rerouting by combining deliveries and new return requests could improve operational efficiency. Moreover, inventory deliveries which are based on sales data in a Vendor Managed Inventory (VMI) framework can benefit from dynamic routing. In this case, routing is organized on a short-term basis and dynamic inventory routing could be applied. Special purpose models can provide considerable help in improving the routing schedule, increasing the precision of the future demand estimations and decreasing the requirement of extra (non-prescheduled) vehicle routes and their respective costs (Wang et al., 2016; Sage, 2013).

Another important aspect in the procedure of e-commerce logistics is the "reverse logistics" as the European law gives to customers the opportunity to return the product within 14 days from the order date. The total return percentage varies per industry. It is not surprising that specialty stores and food have a low return rate. Consumers already know what they are looking for in a specialty store, and fresh products such as food are hard to return most of the time. This is also the case for our interviewees' companies that represent electronics industry. However, in fashion the return rate is much higher than in other industries. The main reason for this is that clothes have different sizes, different fabrics, etcetera. For example, the total return percentage in fashion industry is 16,50%. However, some companies believe it is not a bad thing to have a high return rate and that it is related to the overall company strategy. However, their collection adds complexity in the distribution processes and increases the distribution cost. A major problem that many 3PLs companies also face is related to the reverse logistics and the empty runs. Even if the load factor of a vehicle is high and very satisfying, the same vehicle returns without any goods at the end of the day, and this is translated only to costs (regardless of whether it is paid by the companies themselves or not). Therefore, collaboration in the returns area was recognized as a potential scenario of applying big data solutions.

However, the limited traceability of the orders and the restrictive process of deliveries that do not allow consumers to alter the point of delivery seem to be responsible for the lack of dynamically changing delivery options and processes. In this context, delivery efficiency can be considerably improved by combining routes of multiple supply chain stakeholders (e.g. e-commerce retail players, 3PLs) and by implementing consolidated pick-up and delivery solutions (Tyan et al., 2003; Danloup et al., 2014). Consolidating routes from different companies can help in making more efficient the routing process and also increase the utilization capacity of vehicles. This requires exploiting and consolidating information from multiple data sources. This use case appears to be a novel feature in the practice of collaborative logistics in the context of e-commerce deliveries.

Moreover, companies declare that most of customers' complaints are related to the wide time window delivery that forces then to wait at home until they receive the product. As the time window in home deliveries cannot yet be decreased, retailers investigate the possibility of alternative delivery options. Current delivery options include home deliveries, deliveries at physical stores or couriers' offices and deliveries at pick up points e.g. gas stations. All the interviewees expressed their interest for providing more advanced delivery options at their customers and they mentioned that they have already explored various options e.g. port baggage delivery options. Click and collect service is an alternative delivery method that can offer flexibility and cost reduction and thus increase customers' satisfaction. Xu et al. (2014), argue that pick up points appear to be a novel way for limiting the costs of last mile e-commerce delivery. According to McKinsey global institute report (2013) the future of online grocery in Europe is closely related to successful operation of the click and collect service. This model gives retailers easier entry into the online-grocery space, since it has much less daunting economics than home-delivery service. In the case of e-commerce, adopting a hybrid system of retail stores and click and collect points by adding click and collect points to an existing distribution network on one hand offers an additional service to the consumers, but on the other entails the risk of creating points of unutilized capacity. For these reasons making the right mix of physical retail stores and click and collect points and finding the right location of click and collect points is a strategic decision concerning the distribution network that should be carefully evaluated (Gulati and Garino, 2000).

4.2. Business requirements

Following the approach described in the previous section, we concluded to a set of business requirements that express the needs towards the e-commerce logistics. The general requirements, that were selected by considering the potential benefits and perceived value, are presented below. Our goal is not to have an exhaustive set of business requirements but to investigate how they could be served by big data analytics solutions, propose a big data analytics architecture and discuss any challenges may raise.

Req. 1: To explore how collaboration could be applied in the e-commerce logistics domain in order to address the current challenges and support firms to decrease cost and improve the overall distribution process performance. Collaboration among supply chain stakeholders appeared to be one of the most challenging facts in the context of e-commerce logistics. Last mile delivery challenges, the demand for more responsive supply chain and the need of cost reduction have brought collaborative approaches in the forefront. For example, consolidated pick-up and delivery process by combining routes of multiple supply chain stakeholders (Tyan et al., 2003; Danloup et al., 2014) can contribute in operational efficiency in the last mile cases. Even though the potential collaboration could be the key on delivery cost reduction, companies in e-commerce are reluctant to share data and common vehicles as came out during the interviews and they need to have more evidence about the value of the collaboration.

Req. 2: To identify potential synergies among the e-commerce stakeholders at the reverse logistics. Reverse logistics is considered to be one of the most complicated and costly process in e-commerce logistics according both to literature and industry interview outcomes. Costs and the average percentage of returns vary significantly between different industries (clothes vs electronics) and it reaches up to 30% in the case of fashion industry.

Req. 3: To analyse current distribution processes in order to depict the current distribution patterns and forecast future problems. The abundance of data (routing data, orders, customer data) in association with big data analysis approach can dynamically contribute to identifying patterns of distribution, enhance the visualization of the distribution processes and also support the identification of the problematic processes and the forecasting of failed deliveries and returns, thus improving operational efficiency.

Req. 4: To provide alternative shipping methods at the consumer in order to increase customers' satisfaction and provide lower prices. Alternative last mile delivery methods such as click and collect and pick up points could provide both lower costs to the retailers, 3PL and courier companies additionally to better customer experience and service to the customers. Locating click and collect and pick up points is a critical requirement in the context of last mile optimization.

Req. 5: To enable dynamically changing supply chains that take into consideration various routing and customer preference characteristics in order to decrease logistics costs and to increase customers' satisfaction. Courier companies according to the interviews have identified customers' need to dynamically change the time and the location of the delivery and are already working on future services to this direction. One of the most challenging requirements on the e-Commerce logistics is the real time dynamic route optimization where real time information from different sources (traffic data/sensors/real time events) is used in order to dynamically optimize routing (Pillac et al., 2013; Fabian and Christian, 2012).

Req. 6: To identify users' problems relevant with the logistics procedure, via the analysis of social media data. Social media growth reveals the opportunity to exploit the open data provided through them. Online consumers express their opinion (negative or positive) via comments, either in blogs/forums or in social networking sites. These data can be analysed to provide insights and identify problems that are relevant with the logistics procedure. Algorithms based on text mining and meaning techniques could be used in order to interpret the data and extract the proper information.

4.3. Use case scenarios

In order to further elaborate on the role of Big data analytics on addressing e-commerce logistics challenges, we formulate a set of use cases scenarios by taking into account both the business requirements and the perceived value that could be derived. Below, we shortly describe each one of the use cases and the role that Big data analytics could play on their implementation. Figure 1 presents the use case scenarios, as well as their correlation with the business requirements.

Use Case Scenario 1: End to end information sharing in the supply chain: An integrated view of supply chain towards collaboration of a 3PL and a courier company. The main objective of this scenario is to map the end to end procedure of logistics by combining data provided both by 3PL companies and couriers (delivery data, product data, fleet data, routing data), and thus investigating possibilities for improving distribution procedures. Re-routing by dynamically combining deliveries and new return requests could improve operational efficiency of the distribution processes. Big data technologies can collect data from various sources (e.g. 3PL companies, couriers, open data) in real time and big data analytics (routing algorithms) can provide solutions that can decrease the delivery cost and also improve the efficiency of the distribution process.

Use Case Scenario 2: Shared micro-hubs among online retailers. An alternative way of distribution models in the e-commerce market involves shared micro-hubs. The shared warehouses can work as a joint hub for different e-shops, retailers and logistics companies, serving as the starting point of the last mile customer delivery. The joint efforts result in a flexible and efficient logistics network for all involved partners and can greatly boost (urban) e-commerce logistics by minimizing travelled distances and resulting in time savings and reduction in distribution cost and CO_2 emissions. Data from different online retailers and 3PL companies could be used in this scenario. The location of micro-hubs is critical for enhancing the distribution processes.

Use Case Scenario 3: Shared reverse logistics among online retailers. This use case scenario supports the concept of collaboration among online retailers. The aim of this scenario is to reduce the cost of collecting the returns as well as CO2 emissions by incorporating the distribution flows from customer delivery points and sharing a common vehicle in order to collect them by the common or adjacent delivery points. This scenario will also incorporate the notion of dynamic re-routing in order to provide alternatives where the product is collected when a courier is nearby in order to deliver products.

Use Case Scenario 4: Identification of delivery patterns and problematic issues and future forecasting. The fourth scenario includes the identification of distribution patterns in terms of space and time, of specific e-commerce retailers in order to enhance inventory management and optimize the delivery process. By analysing historical data, seasonal trends and emerging freight flows, behaviour patterns and specific problems regarding the distribution processes (failed deliveries) can be detected. Moreover, this scenario includes accurate demand forecasting. Forecasting the deliveries per area or per season could provide useful information since it can contribute towards optimizing the logistic processes both in terms of network planning such as location of distribution points and warehouses, as well as in terms of operational decisions regarding fleet capacity planning.

Use Case Scenario 5: Shared Click & Collect Points (in terms of space) where customers collect their online orders. With click and collect points, consumers shop online but then travel to the collection point to collect their purchases. Today, the retailers use their own stores (collection points) when they offer Click & Collect services; so there is still a need to have a trip but the consumer saves the picking time. In this use case scenario, retailers will use click and collect points in central locations of the town (e.g. in bus or train stations or post offices) where the customers could collect their orders. The idea about the shared click and collect points is that these points could be small warehouses with lockers containing mainly small packets or special purpose areas in some of the existing shopping centres. The customer's orders could be delivered in off-peak hours by the retailers/couriers to these points from which the customers could collect their orders.

Use Case Scenario 6: Inventory routing in an omni-channel environment. This use case scenario concerns the collaboration between an online supermarket and a brick and mortar supermarket chain. The basic objective is to reduce the transportation costs, by using the physical network of the supermarket chain as intermediate hubs which will be responsible for serving the requests of the local area. The use case will address the problem of choosing the most appropriate -in terms of cost efficiency- physical stores to transform them into hubs and the routing of the inventory between these intermediate nodes and the central hub. This involves the decision for the quantities, the time and the routing of each of these hubs in a total solution to minimize the costs of the whole procedure. Routing data, fleet data, delivery and order data, are involved in this scenario. The aforementioned problem is formulated as an Inventory Routing Problem (IRP).

Use Case Scenario 7: Online consumer insights regarding logistics and delivery processes. The main objective of this scenario is to identify online consumers' problems and preferences, relevant to logistics procedure, extracting information from social media. The data of this use case comprise of text reviews from consumers; thus, it is what is called "unstructured data". Algorithms based on text mining and meaning techniques could be used in order to interpret the data and extract the proper information.

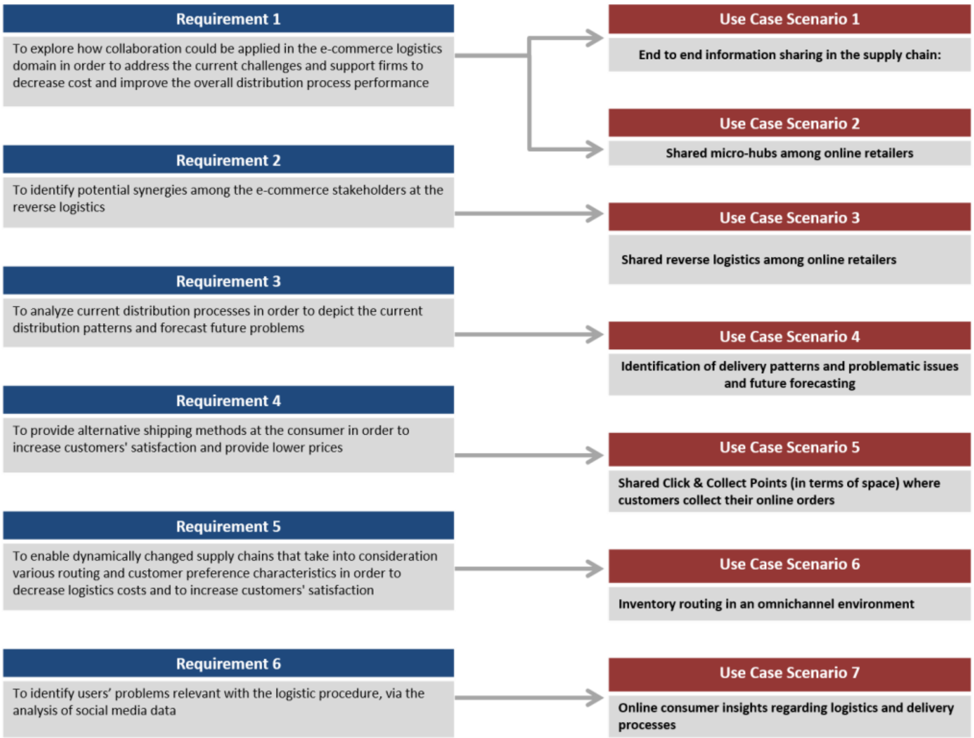

Figure 1- Business requirements and use case scenarios

These specific use case scenarios will be used in order to design the big data analytics artefact.

5. Proposed big data analytics artefact design

Relying on the aforementioned business requirements and the use case scenarios, we identified a set of different technologies and algorithms, which will be used to support the implementation of the use case scenarios. Moreover, we will show how these different technologies and algorithms will be incorporated in the different layers of an overall architecture (Figure 2).

Data sources: The data sources layer includes the external data sources that will be used in order to retrieve data for implementing the various use case scenarios. The data sources include: online retailers' data, 3PLs data, couriers' data and social media data.

Connectors: The main component in this layer is the Core IDS Connector that is designed to exchange data between two or more industrial companies. The Connectors enable a higher security level by combining multiple security mechanisms and the data will only be exchanged between the certified and linked partners. The full infrastructure layout describes the connector in a two parts-component: the company-side connector and the framework-side connector between which the data is exchanged. Before the data transfer, several security and homogenisation services are implemented within the connectors. The framework-side connector is set on top of the real big data storage systems.

Big data storage and computation platform: The big data storage and computation platform is basically a set of network connected servers with regular storage and computation resources, arranged in cluster mode for sharing resources and enhancing performance. The cluster is configured as an Open Source Apache Hadoop Ecosystem. Apache Hadoop is an ecosystem of open source components that fundamentally changes the way enterprises store, process, and analyse data. Unlike traditional systems, Hadoop enables multiple types of analytic workloads to run on the same data, at the same time, at massive scale on industry-standard hardware. Cloudera's open source platform (CDH distribution) is used for deploying the Apache Hadoop Ecosystem.

Data management: This layer is used to organize and model the data stored in the big data infrastructure. This layer will be also responsible for handling any data privacy issues and handling the collaboration aspects of the logistics. The following services are provided:

a. **Annotation/Curation Service:** This service will deal with organizing and annotating data utilized for the scenarios. The service will be developed on top of the Hadoop Cluster, using appropriate tools depending on the available Hadoop distribution. For each dataset (either raw format ingested or processed format), appropriate metadata will be kept in a formalized manner, including dataset description, dataset utilization, version control etc.
b. **Data engineering:** Data engineering service will provide descriptive analytics on each and every data set (data visualization). It will also handle the task of converting each data element to the appropriate equivalent form defined by the data model (data homogenization). Furthermore, data will be cleansed and wrangled to avoid having null values wherever possible. Finally, this service will transform all data in the appropriate forms required by the higher level analytical services. This service will be developed using either Python or Apache Spark, or a combination of both. These tools have been selected having flexibility in mind as well as providing scalability in a distributed manner to handle big data challenges.
c. **Data modelling service:** For the implementation of this service, an appropriate data model needs to be created based on the available data and the use cases. For each element in the data model, specific mappings need to be created in order to be used by different data sources. The data model will be implemented in one of the open source databases included in the infrastructure, and data will be loaded to be utilized by the analytics algorithms.

Big Data Business analytics: This layer includes all the business analytics required for supporting the use case scenarios.

a. **Descriptive analytics and real-time analysis:** A near real time analytics component required that allows understanding and structuring information, identifying hidden patterns and correlations in the data, and inducing knowledge, as well as building learning systems. Its capacity to be flexibly adapted to develop new functionalities through the definition of workflows, which address the complexity of scientific and business applications, new Semantic Models and the rapid development of cloud services gives it a differentiating value. In addition, the way in which data scientists (that work with algorithms and data transformations using a visual interface) and software engineers (that work with the idea of services to be invoked) easily collaborate in order to reduce the “time to market” in the development of complex big data and artificial intelligence applications is another innovative and distinctive feature of Moriarty framework.
b. Sentiment analysis: A data aggregation and sentiment analysis component based on information available in social media. A crawler that collects consumers’ reviews by e-shops, parses user reviews and stores the data in preformatted documents with attributes such as: shop name, shop total votes, date of the review, positive or negative key features (selected form predefined lists), review main text, star rating. The negative and positive reviews are separated and the most used words in these posts are identified. It also performs part-of-speech tagging and keep only nouns, adjectives and adverbs with sentiment (consumer insights), and group user reviews by occurrence of these insights. Then, the previous words are combined (via text mining tools) in order to understand better the way that these words may explain consumer behaviour.
c. **Dynamic Routing:** This component creates optimal routes for both, direct and reverse flows, to optimize the distribution process of different companies, allowing a shared logistics scenario. In order to do this, this

component implements solutions for two principal problems: the vehicle routing problem (VRP) and the vehicle routing problem with time windows (VRPTW). The VRP is a combinatorial optimization and integer programming problem which tries to provide the optimal set of routes for a fleet of vehicles to traverse in order to deliver to a given set of customers. The objective of the VRP is to minimize the total route cost. Therefore, heuristics and genetic algorithms are used due to the size & frequency of real world VRPs they need to solve. The VRPTW is the same problem that VRP with the additional restriction that in VRPTW a time window is associated with each customer, defining an interval wherein the customer has to be supplied. The context is that of delivering goods located at a central depot to customers who have placed orders for such goods. The interval at the depot is called the scheduling horizon.

d. **Deliveries and demand forecasting**: This component implements and configures forecasting methods in order to identify potential failed deliveries in advance. Different forecasting methods are used: statistical, econometric and soft-computing methods. The selection of a method will depend on many factors—the context of the forecast, the relevance and availability of historical data, the degree of accuracy desirable, the time period to be forecast, the cost/ benefit (or value) of the forecast to the company or the time available for making the analysis. A dynamic forecasting method selection behaviour will be configured.
e. **Location optimization:** This component will provide several techniques to analyse the logistics network structure of the partners and to optimize the locations of micro hubs and click and collect points. It supports an established and transparent decision-making support in logistics and offers a broad selection of possibilities, e.g. regarding the extrapolation of shipping volumes, the determination of optimal warehousing sites, route planning etc. It also finds optimal logistics locations based on distances, cost information and shipment volumes. The transports are recorded precisely at SKU-level and are optimised based on real freight rates. General savings in transport costs due to optimised network structures will become visible. Since it considers existing structures and multistage logistics networks, an optimized network approach regarding the location and structure of micro hubs will be found.
f. **Inventory Routing:** The Inventory Routing Problem (IRP) is an operational research problem that has risen relatively recently in the context of Vendor Managed Inventory (VMI). In the classical formulation of the problem, a supplier delivers goods to the customers that are geographically dispersed. The supplier has the full responsibility for not causing stock-out to any of the customers and handles their inventory and stock levels. This means that the supplier is responsible for making the replenishment decisions according to the demands and the imposed constraints of all customers. The problem is considered an NP-complete problem, meaning that it is computationally impossible to examine all possible solutions in order to address the optimal one. A specialized algorithm that use the provided data to examine if it is more efficient to use the physical supermarket network instead of only the central hub and in what extend. The expected benefits in terms of cost can be roughly quantified and the performance can be evaluated towards other metrics such as fleet and vehicle utilization.

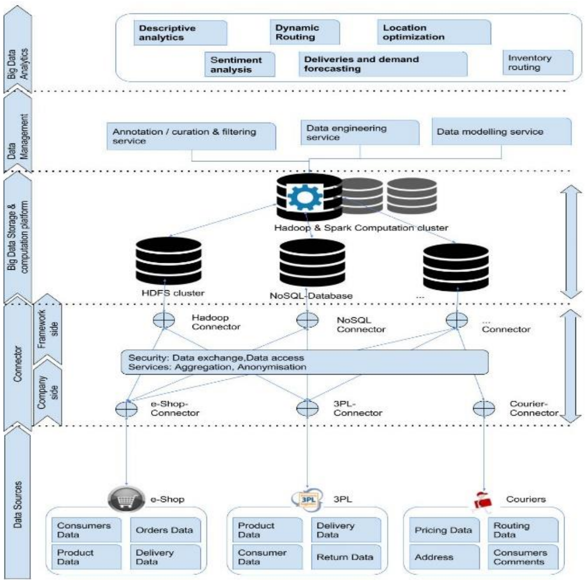

Figure 2- Overall architecture

6. Concluding remarks

This study frames the e-commerce logistics area by identifying a set of business requirements. Then, it provides empirical evidence on how big data analytics serve the identified business requirements, by supporting data integration, decision making and the implementation of seamless logistics practices. An indicative big data analytics architecture is suggested. Next steps include the implementation of the suggested architecture and the respective the use cases using advanced analytics. Implementation challenges will be discussed and special emphasis will be given on illustrating the impact of big data analytics on the transformation of e-commerce logistics.

Acknowledgements

Research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 731932 (TransformingTransport)

References

Addo-Tenkorang, R, and P. T. Helo (2016). Big data applications in operations/supply-chain management: A literature review. Computers & Industrial Engineering 101 (2016) 528–543.

Agatz, N. A., M. Fleischmann and J. A. Van Nunen (2008). E- fulfillment and multi-channel distribution–A review. European Journal of Operational Research 187(2): 339-356.

Akter, S. and Wamba, S.F. (2016). Big data analytics in E-commerce. A systematic review and agenda for future research. Electron Markets, 26, 173-194, 2016.

Aziza, B. (2017). Amazon Buys Whole Foods. Now What? The Story Behind The Story. Available from https://www.forbes.com/sites/ciocentral/2017/06/23/amazon-buys-whole-foods-now-what-the-story- behind-the-story/#26fb066ee898

Bask, Lipponen, and Tinnilä, (2012). E-Commerce Logistics: A Literature Research Review and Topics for Future Research. International Journal of E-Services and Mobile Applications, 4(3), 1- 22, July-September 2012.

Berente, N., and K. Lyytinen (2006). The Iterating Artifact As a Fundamental Construct in Information System Design, *In Proceedings DESRIST,* Claremont, CA: Claremont Graduate University.

Boyer, K. K., Prud'homme, A. M., and Chung, W. (2009). The Last Mile Challenge: Evaluating the Effects of Customer Density and Delivery Window Patterns. *Journal of Business Logistics, 30(1),* 185–201. https://doi.org/10.1002/j.2158-1592.2009.tb00104.x

Castiñeira, R, and Metzger, A. (2018). The TransformingTransport Project – Mobility Meets Big Data. In Proceedings of 7th Transport Research Arena TRA 2018, April 16-19, 2018, Vienna, Austria

Chen, H., Chiang, R. H. L., and Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36(4),* 1165–1188.

Danloup et al., 2014. Improving sustainability performance through collaborative food distribution International Conference on Green Supply Chain, France.

Du, T. C., Li, E. Y., and Chou, D. (2005). Dynamic vehicle routing for online B2C delivery. *The international journal of Management Science* **(*33*)**, 33–45. doi:10.1016/j.omega.2004.03.005

E-commerce Europe. (2016). European B2C E-commerce Report - Facts, Figures, Infographic & Trends of 2015 and the 2016 Forecast of the European B2C E-commerce Market of Goods and Services. Retrieved from https://www.ecommerce-europe.eu/app/uploads/2016/07/European-B2C-E-commerce-Report-2016-Light- Version-FINAL.pdf

Esper, T. L., T. D. Jensen, F. L. Turnipseed and S. Burton (2003). "The last mile: an examination of effects of online retail delivery strategies on consumers." Journal of Business Logistics 24(2): 177-203.

Fabian, J. and Christian, D. (2012). Vehicle routing for attended home delivery in city logistics. *Procedia: Social and Behavioural Science*s (**39**), 622–632. doi:10.1016/j.sbspro.2012.03.135

Gulati and Garino (2000). Get the Right Mix of Bricks and Clicks. Harvard business review, May- Lune 2000 issue.

Hevner, A. R., March, S. T., Park, J. and Ram, S. (2004). Design Science in Information Systems Research. MIS Quarterly, 28(1), 75–105.

Hjort, K., and Lantz, B. (2016). The impact of returns policies on profitability: A fashion e-commerce case. *Journal of Business Research, 69(11),* 4980–4985. https://doi.org/10.1016/j.jbusres.2016.04.064

Jeseke, M., Grüner, M., and Weiß, F. (2013). Big data in logistics: A DHL perspective on how to move beyond the hype. White Paper; DHL Customer solutions and innovation, presented by Martine Wegner, Vice President Solutions & Innovation, 53844 Troisdorf, Germany.

Jones, D. and Gregor, S. (2007). The anatomy of a design theory. Journal of the Association for Information Systems, 8(5), 1.

Koutsabasis, P., Stavrakis, M., Viorres, N., Darzentas, J. S., Spyrou, T., and Darzentas, J. (2008). A descriptive reference framework for the personalisation of e-business applications. Electronic Commerce Research, 8, 173–192.

McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D., (2012). Big data. The management revolution. Harvard Business Review, 90(10), 61–67 McKinsey global institute report (2013). Perspectives on retail and consumer goods

Mehra, G., (2013). 6 uses of big data for online retailers, Practical Ecommerce. Available at: http://www.practicalecommerce.com/ articles/3960-6-Uses-of-Big-Data-for-Online-Retailers (Accessed 2nd of March, 2016).

Miller, G., (2013). 6 ways To use "big data" To increase operating mar- gins By 60 %. Available at: http://upstreamcommerce.com/blog/ 2012/04/11/6-ways-big-data-increase-operating-margins-60-part-2 (Accessed 2nd of March, 16).

Morganti, E., Seidel, S., Blanquart, C., Dablanc, L., and Lenz, B. (2014). The Impact of E-commerce on Final Deliveries: Alternative Parcel Delivery Services in France and Germany. *Transportation Research Procedia, 4(Supplement C),* 178–190. https://doi.org/10.1016/j.trpro.2014.11.014

Moroz, M., and Polkowski, Z. (2016). The Last Mile Issue and Urban Logistics: Choosing Parcel Machines in the Context of the Ecological Attitudes of the Y Generation Consumers Purchasing Online. *Transportation Research Procedia, 16(Supplement C),* 378–393. https://doi.org/10.1016/j.trpro.2016.11.036

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24(3), 45–77.

Pillac,V., Gendreau, M., Guéret, C., and Medaglia, A.L., (2013). A review of dynamic vehicle routing problems. *European Journal of Operational Research 225*(**1**) 1-11. http://dx.doi.org/10.1016/j.ejor.2012.08.015.

Punakivi, M., Yrjo, H., and Holmstro, J. (2001). Solving the last mile issue: reception box or delivery box? International Journal of Physical. Distribution & Logistics. Management 31(6), 427–439

Sage, (2013). Better inventory management: Big Challenges, Big Data, Emerging So- lutions. 〈http://na.sage.com/media/site/erp/responsive/resources/Sage-ERP-Bet ter-Inventory- Management-wp.pdf〉 (accessed 20.04.15).

Sharma, R., Mithas, S., and Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organizations. *European Journal of Information Systems, 23*, 433–441.

Swaminathan, S., 2012. The Effects of Big Data on the Logistics Industry. Available at http:// www.oracle.com/us/corporate/profit/archives/opinion/021512-sswaminathan- 1523937.html〉

Tyan, J. C., Wang, F.-K., and Duc, T. C. (2003). An evaluation of freight consolidation policies in global third party logistics. Omega, 31(1), 55–62. doi:10.1016/S0305-0483(02)00094-4

Wamba, S., Akter, S., Edwards, A., Chopin, G., and Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 165(Supplement C),* 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031

Wang, G., Gunasekaran, A., Ngai, E. W. T., and Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics, 176,* 98–110.http://doi.org/10.1016/j.ijpe.2016.03.014

Xu J., Jiang L., and Wang S. (2014) Construction of Pick-Up Points in China E-commerce Logistics. In: Zhong S. (eds) Proceedings of the 2012 International Conference on Cybernetics and Informatics. Lecture Notes in Electrical Engineering, vol 163. Springer, New York, NY