DISCLAIMER

This deliverable has been submitted but has not been approved by the EC yet

"Enhanced data management techniques for real-time logistics planning and scheduling"

Deliverable D1.2: LOGISTAR technologies market research

Dissemination level:
☒ Public ☐ Confidential, only for members of the consortium (including the Commission Services)

Version number: 1.0

Submission deadline: 30/11/2018

www.logistar-project.eu

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 769142.

DOCUMENT INFORMATION

Authors

Name	Organisation
Andrew Palmer	Preston Solutions Ltd
Gunter Klein	dbh Logistics IT AG
Gaetano Formisano	Genegis GI
Artem Revenko	Semantic Web Company GmbH
Nenad Gligoric	DunavNET

Reviewers

Name	Organisation
Jason Cronin	Preston Solutions Ltd
Bastien Pietropaoli	UCC
Enrique Onieva	University of Deusto
Naia Merino	University of Deusto

Document control

Version	Date	Comment
V0.1	19/11/2018	First version of deliverable 1.2
V0.2	22/11/2018	Included revisions in sections 3 and 5
V0.3	30/11/2018	Included revisions from UCC & Deusto

Document approval

Version	Date	Partners
V1.0	03/12/2018	All partners

BASIC PROJECT INFORMATION

Horizon 2020 programme

H2020 - Mobility for Growth- 5-2-2017. Innovative ICT solutions for future logistics operations

Grant Agreement No. 769142

TABLE OF CONTENTS

Executive summary 6
1. Introduction 7
2. Cloud computing 8
3. Internet of Things (IoT) 9
4. Event detection 14
5. Broad data in transport and logistics 24
6. Logical infrastructures for big data analytics - modern ETL approaches 29
7. Geographic information systems 35
8. Conclusions 48
List of abbreviations and acronyms 49

LIST OF FIGURES

Figure 1: Crane collapsed in Bremerhaven 17
Figure 2: Example Ransom ware detection 18
Figure 3: CSD ORBCOM 20
Figure 4: Traffic communication (example) 22
Figure 5: ETA (Estimated Time of Arrival) 22
Figure 6: Stakeholder groups (fragmentary) 25
Figure 7: UnifiedViews architecture 32
Figure 8: System kernel 34
Figure 9: Logistics ontologies overview 34
Figure 10: Transport resource ontology diagram 35
Figure 11: Multi-tier architecture 37
Figure 12: Typical Geoportal system 39
Figure 13: ROOTS Green Logistics - ARCHITECTURE 40
Figure 14: ROOTS Green Logistics – Interoperability 41
Figure 15: ROOTS Green Logistics - Planning Interface 42
Figure 16: Real-time monitoring of the current position of the vehicle 43
Figure 17: Monitoring of deliveries in real-time 43
Figure 18: Estimate arrival time for customers to be served 43
Figure 19: Real-time route recalculation and re-planning in case of logistical difficulties 44
Figure 20: Analysis of time reliability 44
Figure 21: Real-time management of fleet delays 45
Figure 22: Delay details 45
Figure 23: Real-time temperature monitoring system 46
Figure 24: Real-time monitoring user interface 46
Figure 25: Report example 47

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

LIST OF TABLES

Table 1: Fundamental event classification 15
Table 2: Aspects of events 16
Table 3: Incident classification Bremerhaven gantry cranes 17
Table 4: Incident classification earthquake New Zealand 18
Table 5: Incident classification ransomware NotPetya 19

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Executive summary

This report assesses the state of the art for a number of technologies and techniques to be used in the development of the LOGISTAR system. These include cloud computing, the use of IoT devices and systems, event detection, broad data applied to transport & logistics, logical infrastructures for big data analytics and geographical information modules. It is discussed from the perspective of principles, concepts, best practices, examples and case studies.

Cloud computing is now ubiquitous for hosting and delivering services over the internet. The discussion in this report highlights the main concepts and some of the logistics related implementations that have taken place.

As with cloud computing, the internet of things (IoT) is now more prevalent, and organisations are starting to implement increasingly complex systems using a range of devices such as RFID tags, smartphones and embedded sensors. The use of IoT devices, with the supporting hardware and software, provides unique real-time tracking opportunities for freight transport. The report discusses the advantages and challenges associated with IoT including security and privacy issues. In addition, it is crucial that the issue of interoperability and standards are addressed otherwise there is a risk of fragmentation, duplication and competition.

The section on event detection covers the impact of advancing globalization and the expanding supply chain, as well as taking into account aspects such as climate change and the associated changes in weather conditions or the increase in freight transport movements. Forward-looking planning based on forecast data or precise planning corrections due to sudden events have a direct influence on the smooth transportation of goods. It first deals with a methodology for dividing events into groups and then classifies a variety of events relevant to transportation. As not all events and the technical possibilities for their early detection can be dealt with, the second section focuses primarily on the events with which stakeholders are confronted on a daily basis.

IoT, event detection and big/broad data are inextricably linked. IoT and event detection enable vast amounts of information to be collected and analysed computationally to reveal patterns, trends, and associations through data mining and analysis, to provide insights into how supply chain operations can be improved. Big data uses a variety of statistical techniques and computer algorithms, such as artificial intelligence, to identify industry trends and business information from large databases. This report discusses various ETL tools and technologies for capturing data to enable this information to be analysed

The purpose of the section on Geographic Information Systems (GIS) is to define the basic principles and to describe an example of a web-based cartography system used to resolve logistics and workforce management questions. GIS is a combination of hardware, software, and data which is the core of any GIS. This section introduces geodatabases which have a locational set of information and are formed of raster and vector layers. The methods of visualising, manipulating and monitoring of geospatial data are described.

1. Introduction

The main aim of the LOGISTAR project is to allow effective planning and optimising of transport operations in the supply chain by taking advantage of vertical and horizontal collaboration among different sectors and companies and using increasingly real-time data gathered from the interconnected environment such as Internet of things (IoT) devices, smartphones, on-board units and open data. To achieve this, a real-time decision and visualisation tool of freight transport will be developed using advanced algorithms, big data analytics and artificial intelligence which will deliver key information and services to the various agents involved in the supply chain such as freight transport operators and their clients.

This aim will be achieved by:

- Identifying logistics related open data sources and harmonize this data together with the other closed sources (i.e. IoT devices and company data)
- Increasing the accuracy planning of logistics operations by applying artificial intelligence techniques for timing predictions and learning preferences of logistics chain participants
- Ensuring a seamless flow of the operations in the supply chain making use of machine learning techniques for identifying potential disrupting events and taking relevant actions to modify any required reconfigurations
- Making the best use of the available resources and provide the best possibilities for horizontal collaboration among logistics agents applying optimisation techniques to transhipment planning and scheduling in hubs and freight transport networks
- Allowing negotiation among different agents involved in the supply chain taking into account any constraints arising in real-time, making use of distributed constraint satisfaction techniques

This will require the application of a number of technologies and techniques including cloud computing, the use of IoT devices and systems, event detection, broad data applied to transport & logistics, logical infrastructures for big data analytics and geographical information modules.

There are many developments taking place in transport and logistics, particularly in transport management systems (TMS), with telematics and remote sensing of trucks and other transport modes being key themes. Added to this is functionality associated with big data and broad data, plus advanced algorithms, to obtain improved routing taking into account weather conditions and other external factors. Collaboration and communication are essential for current TMS applications within cloud-based technology.

This deliverable explores the state of the art in these technologies from the perspective of principles, concepts, best practices, examples and case studies. The following sections cover cloud computing, IoT devices, event detection, broad data, big data analytics and geographical information modules.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

2. Cloud computing

Cloud computing has developed rapidly since it was first popularised by Amazon.com in 2006, through its Amazon Web Services subsidiary, and its Elastic Compute Cloud product. A key driver for these changes has come about because of the rise in social media which is influencing expectations about business applications, and the worldwide growth in networks of logistics companies. The “Amazon effect’ has disrupted the industry because companies that are unable to provide real-time updates on request risk falling behind and losing customer confidence.

An increasing number of transport management systems are now cloud-based, having moved from monolithic applications to more granular cloud-based micro services delivered via software as a service (SaaS) models. An important technology in this context are APIs (Application Programming Interfaces) which makes it possible to quickly access specific information without having to build, populate or maintain a whole data pool individually. With cloud-based systems supply chain companies can get real-time access to information throughout the logistics ecosystem. Consequently, the amount of available data is constantly increasing making big data applications very attractive. making it a key enabler of a virtual ‘control tower’, providing 360-degree management dashboards. Cloud computing storage technology can improve the efficiency of big data collection and processing. In a collaborative environment, TMSs can be integrated across many different partner systems in cloud computing which means that they can handle a much larger volume of shipments, locations and vehicles, and optimise supply chain performance.

Cloud computing enables rapid, efficient, and flexible access to IT services for innovative supply chain solutions. Currently, more than 50% of logistics providers use cloud-based services and a further 20% are planning to do so in the near future[1]. In the future, open and web-based APIs will form the basis of modular on-demand cloud logistics services, replacing outdated, legacy communication systems (such as EDIs).

Logistics portals provide flexible and configurable systems that can be integrated into supply chain processes. For example, many cloud-based transport management systems can process orders, invoicing, load building and scheduling, and some also offer track-and-trace functionality. These systems aren’t restricted to large companies. Small and medium-sized logistics providers can take advantage of software as a service (SaaS) or logistics as a service (LaaS) options on a pay as you use basis. This avoids them having to invest in a fixed IT infrastructure. EC research projects Aeolix & Selis, and an earlier development MixMoveMatch, are developing, or have developed, cloud-based portals for logistics services.

Cloud-based logistics systems create a virtual version of information by moving all supply chain processes into the cloud. Supply chains have become increasingly complex and fragmented due to globalisation and customised products and services. Consequently, shippers and logistics service providers may have to deal with a range of transactions between different parties, typically using different warehouse and transport management systems. A cloud-based system enables this information to be coordinated into one integrated view.

[1] https://www.logistik-heute.de/Logistik-News-Logistik-Nachrichten/Markt-News/15213/Cloud-Loesungen-werden-fuer-die-Logistik-immer-wichtiger-Studie-Siegeszug-de

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Bosch has developed a new connectivity platform which serves to connect commercial vehicles throughout their service life, providing the technological foundation for cloud-based services such as predictive diagnostics and over-the-air software updates. The platform has two main components: the basic software module is the secure communication interface between the vehicle, the cloud and the services, while the data management module enables commercial vehicle manufacturers or fleet managers to organise vehicle data, analyse it and keep vehicle software continuously updated. A digital key connects trucks and the smartphone app via the cloud. Dispatchers or fleet managers use the app to assign a truck to a driver for a particular route. It then generates a personal, secure digital key and sends it via the cloud to the truck and to the driver's smartphone. As the driver approaches the assigned truck, the sensors installed in the truck detect the smartphone via a wireless connection. The vehicle doors will open only if the key on the phone 'fits' the digital lock in the vehicle. These sensors can also tell when someone is in the driver's seat, and the engine fires up as soon as the start-stop button is pressed. When the driver gets out of the truck at the end of the journey, the system detects this and automatically locks the doors.

Volvo Trucks is also developing cloud-based technology, with a focus on the safety benefits the technology can bring. It not only allows trucks to talk to each other and to operators, but also to Volvo passenger cars to collect real-time data around traffic conditions and potential hazards.

Other examples include:

"Shipwire Enterprise Logistics Platform, Westfracht Spezialverkehre: LaaS (logisticsas-a-service.de), LogFire Cloud Solutions, Transporeon – logistics services from the cloud, Salesforce – cloud-based services for small to large enterprises, Freightly – real-time cloud-based system covering all logistics processes for transportation."[2]

3. Internet of Things (IoT)

The term "Internet of Things" was first introduced in 1999 by the MIT Automatic Identification Research Cente to describe objects that contained radio frequency ID (RFID) tags. According to Thomas (2018)[3] "A typical IoT device will either have a sensor so that it can measure something, or an actuator so that it can change something in the physical world, or both". Some IoT devices are increasingly seen in small electronics such as those found in homes or worn on the wrist, others are very large national scale applications such as a water diversion project in China involving 3 canals over 1,000km long. Manually monitoring the canals, particularly its two tunnels, would be extremely difficult, so a network of over 100,000 IoT devices have been installed in a 1,257km stretch, scanning the canal for structural weaknesses, testing water quality and flow rates and watching for intruders.

There are many IoT systems in use[2,3] providing significant benefits such as:

- sensors in Rolls Royce aircraft engines streaming back usage and condition of the engines for optimising maintenance
- data from connected vehicles and consumer smartphones to measure traffic flow in real-time and fed back to satnavs for route guidance and build an intelligent transport system

---

[2] DHL Trend Radar (2018)
[3] Thomas, M. (2018) "The Internet of Things", Gresham College

- connected CCTV cameras and automated number plate recognition which can track vehicles in real-time
- enabling retailers to manage stock levels and detect shoplifting
- DHL/Cisco IoT warehouse project
- DHL Smart Sensor Solutions
- Agheera – the IoT platform for logistics
- DHL Clever Lock – enabling unattended home deliveries with smart locks
- Kwik – IoT push buttons enabling automated ordering of products
- Physical Internet Initiative – smart container management with real-time monitoring
- BeeBright – start-up focusing on smart lighting for larger spaces such as warehouses
- Volvo Maintenance on Demand

Over the last two decades, supply chains have undergone significant change. What was previously led by human resources, now is shifted to advance planning processes, such as demand and operation planning leaving logistics mostly outsourced to a third-party logistics provider. An Internet of Things in logistic is used to track assets to optimize supply chain (tracking deliveries from vendor to destination, fleet management to monitor assets and mitigate risks), monitor sensitive goods and assets conditions to avoid damage and loss (tagging of goods to monitor temperature, humidity, vibration, etc). In LOGISTAR it is mandatory to provide efficient data collection and pre-pre-processing in WP2 to enable efficient implementation of data exchange modules and deployment scenarios in WP6, modules like the Decision Proposal Calculation Module, the LOGISTAR Dashboard, the Geographic Information Module, and the Messaging and Complex Event Detection Module.

Further to this, transport solutions incorporating IoT can increase transparency and integrity in the supply chain by, for instance, in-cab telematics, which can collect data on movements and idle time to ensure fleet asset utilisation is maximised. This feature could also provide dynamic route planning and optimisation.

IoT presents a huge opportunity in logistics[4]. However, until now only a few IoT applications in logistics have experienced widespread adoption, due to the total cost of deployment, security concerns, and an absence of standards in the fragmented logistics industry. With the global connected logistics market estimated to grow substantially, breakthroughs in the development of low-cost IoT networks and continually falling sensor prices, will enable large-scale IoT deployments in logistics. Furthermore, early adopters will have opportunities to develop new, IoT-based business models.

Delivering a better service to customers is a key initiative for many companies. Using IoT devices can bring an improved understanding of customers, plus reduced costs and increased efficiencies, and the development of new business models.

[4] http://www.cisco.com/c/dam/en_us/services/portfolio/consulting-services/documents/consulting-services-capturing-ioe-value-aag.pdf

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

3.1. Development and application of IoT in freight transport and logistics

The architecture of IoT can be divided into four layers: the perception layer, the gateway and network layers, the management service layer and the application layer[5]. The primary task of the perception layer is to collect and transmit data. The role of the gateway and network layer is to connect objects or things and allow them to share and exchange information[6]. The management service layer relies on middleware technology, which provide services and applications, with features that are seamlessly integrated into the IoT[7]. It is responsible for information analysis, security control, process modelling and device management. The application layer is the ultimate goal of IoT technology. In this layer, the collected and converted data is retained, processed by some techniques for planning, and objects or things are managed and controlled[8].

The concept of IoT has expanded the carrier objects of the traditional Internet, from computers, mobile phones and other devices to items commonly found in daily production and life such as forklifts, pallets, mechanical equipment, and wearable devices[9].

IoT covers a variety of technologies, including LAN, wireless Bluetooth, wide-area connectivity, wired connections, RFID technology and other technical means to collect data on each device[10]. Currently, the combination of IoT technology and warehouse management system has helped many companies to manage warehouses more efficiently[11].

According to research by Zeres Technologies, enterprise IoT deployments have increased by 333% since 2012.

The Port of Hamburg has installed more than 300 sensors to monitor traffic conditions in the port area and the wear and tear of bridges[6]. Through the "smartPORT" program, based on IoT, the Port of Hamburg has succeeded in improving operational efficiency and reducing the impact of port transportation on the lives of nearby residents.

The logistics company UPS has installed sensors on each of their vans. Using IoT technology, UPS can monitor vehicle parts and identify parts that need to be replaced in real-time. Previously, UPS replaced vehicle parts every two or three years. With optimisation techniques linked to IoT technology, UPS has saved several million dollars each year[12].

IoT technology is playing a role in the supply chain of fresh produce[13]. The application of IoT enables the monitoring of production, transportation, distribution, sales testing, traceability and other supply

---

[5] Dweekat, A.J., Hwang, G. and Park, J. (2017b) 'A supply chain performance measurement approach using the internet of things: Toward more practical SCPMS', Industrial Management and Data Systems, 117(2), pp. 267–286.
[6] Zhou, L., Chong, A.Y.L. and Ngai, E.W.T. (2015) 'Supply chain management in the era of the internet of things', International Journal of Production Economics, 159, pp. 1–3.
[7] Luo, H., Zhu, M., Ye, S., Hou, H., Chen, Y. and Bulysheva, L. (2016) 'An intelligent tracking system based on internet of things for the cold chain', Internet Research, 26(2), pp. 435–445.
[8] Dweekat, A.J., Hwang, G. and Park, J. (2017a) 'A supply chain performance measurement approach using the internet of things', Industrial Management & Data Systems, 117(2) Wembley: Emerald Group Publishing Limited, pp. 267–286.
[9] Macaulay, J., Buckalew, L. and Chung, G. (2015) 'Internet of Things in Logistics', DHL Trend Research, 1(1), pp. 1–27.
[10] Rezaei, M., Akbarpour Shirazi, M. and Karimi, B. (2017) 'IoT-based framework for performance measurement', Industrial Management & Data Systems, 117(4) Wembley: Emerald Group Publishing Limited, pp. 688–712.
[11] Lee, C.K.M., Lv, Y., Ng, K.K.H., Ho, W. and Choy, K.L. (2017) 'Design and application of Internet of things-based warehouse management system for smart logistics', International Journal of Production Research, 7543, pp. 1–16.
[12] Li, B. and Li, Y. (2017) 'Internet of things drives supply chain innovation : A research framework', *The International Journal of Organisational Innovation*, 9(January 2017), pp. 71–93.
[13] Yan, B., Wu, X., Ye, B. and Zhang, Y. (2017) 'Three-level supply chain coordination of fresh agricultural products in the Internet of Things', *Industrial Management & Data Systems*, 117(9) Wembley: Emerald Group Publishing Limited, pp. 1842–1865

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

chain activities in the fresh produce sector[14]. This supports the track and trace requirements of government regulatory agencies whilst ensuring the efficiency of the supply chain and all aspects of agricultural product quality inspection[10]

Sony have started using a mobile IoT device in trailers to detect light. This is to check when a door is open, or someone has broken through the trailer roof, for instance. An event is triggered if light is detected at an unexpected time to try and catch thieves. Sony's logistics partners then return the mobile IoT device on completion of a delivery.

Battery life of an IoT device is important. The smaller the device, the smaller the battery, so the frequency of event information being transmitted may need to be longer than a larger IoT device, in order to save battery life.

3.2. Advantages of IoT technology

The application of IoT can improve logistics operational efficiency, security, and customer satisfaction[15]. Through IoT technology, operators can monitor real-time assets in the entire logistics value chain and keep abreast of the status of packages and delivery personnel[9]. By analysing the entire process, IoT technology can help companies find ways to optimise their processes, thereby reducing costs and identifying potential business opportunities[16].

IoT technology links goods, pallets, and operating equipment in the warehouse[13]. These items improve the visibility in the entire warehouse by feeding back information about their work tasks, work conditions, work locations, and more[17]. With real-time data feedback, IoT technology can help companies set more precise supply chain performance management standards[8].

IoT technology puts employees in the warehouse in a more comfortable and safer working environment[18] by using pressure sensors to identify when a forklift is overloaded, or a load is unstable[19]. By collecting employee mobility information, warehouse managers can change walking pathways or operational processes to avoid accidents.

Intelligent transportation systems (ITS) also benefit from the use of IoT devices. ITS aims to provide innovative services relating to different modes of transport and traffic management and enable users to be better informed and make safer, more coordinated, and 'smarter' use of transport networks. It is particularly useful for predicting road travel times. In Shi et al, 2017[20] an estimation of travel time was based on two types of road sensors: point detectors (e.g. loop detectors) and interval detectors (e.g. automated vehicle identification systems). To achieve this, they merge the two travel time distribution estimates using Dempster-Shafer theory and then

---

[14] Li, Z., Liu, G., Liu, L., Lai, X. and Xu, G. (2017) 'IoT-based tracking and tracing platform for prepackaged food supply chain', *Industrial Management & Data Systems*, 117(9) Wembley: Emerald Group Publishing Limited, pp. 1906–1916
[15] Lee, I. and Lee, K. (2015) 'The Internet of Things (IoT): Applications, investments, and challenges for enterprises', *Business Horizons*, 58(4) 'Kelley School of Business, Indiana University', pp. 431–440.
[16] Lee, C.K.M., Lv, Y., Ng, K.K.H., Ho, W. and Choy, K.L. (2017) 'Design and application of Internet of things-based warehouse management system for smart logistics', *International Journal of Production Research*, 7543, pp. 1–16.
[17] Chung, G., Gesing, B., Chaturvedi, K. and Bodenbenner, P. (2018) *Logistics Trend Radar*.
Cio, J. and Apr, F. (2016) '*What Is Blockchain in Simple Terms and How Does It Work?*', , pp. 2016–2018.
[18] Haddud, A., DeSouza, A., Khare, A. and Lee, H. (2017) 'Examining potential benefits and challenges associated with the Internet of Things integration in supply chains', *Journal of Manufacturing Technology Management*, 28(8), pp. 1055–1085
[19] Hopkins, J. and Hawking, P. (2018a) 'Big Data Analytics and IoT in logistics: a case study', *The International Journal of Logistics Management*, 29(2), pp. 575–591
[20] Shi, C., Chen, B. & Li, Q., 2017. Estimation of Travel Time Distributions in Urban Road Networks Using Low-Frequency Floating Car Data. *ISPRS International Journal of Geo-Information* , 6(8), p.253.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

impute travel times for neighbouring links. Their method outperforms the estimates computed from only one of the two types of sensors.

At the same time, IoT devices could help vehicles reduce downtime through forecasting to achieve higher asset utilisation, enhance the customer experience[21], and reduce the cost of product return[22].

IoT technology could be used to support automated warehouse operations[23]. Real-time feedback of information could allow devices to adjust their working state autonomously depending on conditions. When combined with artificial intelligence and data analysis technology, this may eventually achieve a fully autonomous automated warehouse[24].

3.3. Challenges of implementing IoT

The global connected logistics market is estimated to substantially grow over the next 5 years[25]. Therefore, issues related to how to represent, store, interconnect, search, and organize information generated by IoT devices will become very challenging. The technology of IoT has developed into a multi-functional network including tags, sensors, actuators, mobile phones, etc., and with the logistics industry being fragmented and complex the development of IoT standards is essential. In addition, one of the biggest concerns are related to the management of trust, privacy and security of all the exchanged data[17].

Many companies are not aware of the potential benefits that IoT technology can bring to the enterprise, plus there is a shortage of people with the necessary digital skills, which holds back the implementation of the IoT technology[18].

The application of IoT technology in enterprises requires integration at the technical level, not only hardware support but also data analysis software[11]. Enterprises don't appear ready to accept this technological innovation and may also lack corresponding networks and data services[26].

In Thomas (2018)[3] "There are also problems of complexity and the unintended consequences of newly introduced IoT devices interacting with existing IoT devices and forming unintended systems with unforeseen properties and feedback loops."

IoT devices also need networking capabilities that suit their environment, perhaps by using nearby IoT devices to relay data packets[26].

One of the biggest challenges of the IoT is processing large amounts of data and the extraction of qualitative information. To be able to efficiently process the data, the collection and preparation of big data for the machine learning can be done. Another point is that most traditional machine-learning based systems are designed with the assumption that all the collected data would be completely loaded into memory for centralized processing. However, as the data keeps getting bigger and

---

[21] Zhu, D. (2018) 'IOT and big data based cooperative logistical delivery scheduling method and cloud robot system', Future Generation Computer Systems, 86 Elsevier B.V., pp. 709–715.
[22] Abdel-Basset, M., Manogaran, G. and Mohamed, M. (2018) 'Internet of Things (IoT) and its impact on supply chain: A framework for building smart, secure and efficient systems', Future Generation Computer Systems, 86 Elsevier B.V., pp. 614–628.
[23] Addo-Tenkorang, R. and Helo, P.T. (2016) 'Big data applications in operations/supply-chain management: A literature review', *Computers and Industrial Engineering*, 101 Elsevier Ltd, pp. 528–543.
[24] Sreedhar, S. (2017) 'Four Ways the Internet of Things Will Transform the Supply Chain', *Material Handling & Logistics*, Cleveland: Informa
[25]www.prnewswire.com/news-releases/global-connected-logistics-market-expected-to-reach-27722-million-by-2023---allied-market-research-675151133.html
[26] www.researchgate.net/publication/270509690_A_Survey_on_Wireless_Mesh_Network_and_its_Challenges_at_the_Transport_Layer

bigger, the existing machine learning techniques encounter great difficulties when they are required to handle the unprecedented volume of data. There is a great need to develop efficient and intelligent learning methods to cope with future data processing demands. These issues are discussed in more detail later on in this paper.

4. Event detection

There are two main categories of events, those that are expected/planned and those that are unexpected. The following section discusses the latter of these categories.

4.1. Event-Classification

This section considers possible division into event groups and event classification. Since not all events and the technical possibilities for their early detection can be dealt with, the focus in the latter part of this section is primarily on the events with which the stakeholders are confronted on an almost daily basis.

Before considering the technical possibilities with regard to event detection across the entire supply chain, it’s important to identify and classify possible events that could have a direct or indirect effect on the smooth running of freight transport. Furthermore, it’s important to specify various parameters within the classifications for a better understanding of which event groups could affect the transportation process.

This classification reveals which events may be foreseeable at the technical level so that preventive/corrective measures can be taken before the event occurs ("act instead of react") or which events are unforeseeable and can therefore only be recorded at the technical level and made available to relevant systems after the event has occurred.

In addition, a distinction must be made between events that relate indirectly to transport but directly to the goods or loading units (e.g. containers) to be transported and events that relate directly to transport. In the context of the classification, however, it should be noted that the effects of a prolonged traffic jam on the highway can be just as significant with regard to the transport of goods as, for example, a temporary port closure extending over several days.

There are a huge number of events that might occur in a road or rail freight transport operation, examples of which are below. These are typical but seldom major:

- Traffic jams
- Traffic accidents
- Missed time windows for pick-up or delivery
- Unavailability of promised trucks or drivers
- Documents (e.g. customs) which arrive too late and block the process
- Unforeseen waiting times at warehouses or terminals
- Bad weather
- Theft (seldom in Europe; but e.g. in our tobacco transports in Russia)
- Empty charter trucks or re-use containers which are unclean, so the customer refuses to load his goods in it
- Freight damage

4.1.1. Classes of Incidents

This section provides a classification of incidents. An incident in this environment might be described as an event which influences the transport of goods in a negative way.

Incident class	Description / Example	Remark
Environmental	Storm	
	Earthquake	> Southern Europe
	Ice	> North of Europe
	High tide, Spring tide, storm surge	> North of Europe
Technical	Wear and material fatigue	
	Electrical or electronic malfunction	
Intended disruption	Sabotage	
	Terrorism and other severe (?) threats	
	Crime	
Health threat	Illness	
	Epidemic	
	Pandemic	
Peripheral reasons	Strike	
	Political or institution based influence	
	Natural caused delays	
	Other disturbances	

Table 1: Fundamental event classification

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

4.1.2. Aspects of incidents

In order to relate the technical complexity of the digitized event detection in relation to the event effects and the duration of the event, it is necessary to identify further parameters in the context of the classification.

Aspect	Description Example	Remark
Predictability	High predictability	
	Medium predictability	
	Low predictability	
Commencement	Immediate impact	
	Medium term impact	
	Long term impact	
Duration of an incident	Short period	
	Medium period	
	Long duration	
Measure of economic damage	Low	
	Medium	
	High	
	Fatal	
Impact	Direct impact	
	Indirect impact	
Reduction of impact results	Manipulable	
	Partly manipulable	
	Unmanipulable	

Table 2: Aspects of events

4.2. Examples of incidents and classification

To illustrate the extent to which events have a direct or indirect influence on a smooth transport process, some examples are listed below and an appropriate classification is made. Here it becomes clear that a timely and target-oriented event detection is not possible at all times or only very difficult for a large number of events.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Gantry Crane collapsed in Bremerhaven, Germany

Incident description

Figure 1: Crane collapsed in Bremerhaven[27]

On 14th May 2015, the jib of a gantry crane including the cabin broke off and dropped onto the container vessel Maersk Karachi killing the crane operator. During subsequent welding repairs the vessel caught fire.[28] German TÜV (Technical Inspection Association) assessed the accident and found a pre-induced crack of the anchoring causing the incident. The reason for the accident was found about a year after the accident.[29] Eight further gantry cranes of the same type were deactivated for security reasons but were inspected after 8 months and became operational[30].
During this period, ships were handled at MSC's neighbouring terminal or diverted to Germany's first deep-water port Jade-Weser-Port in Wilhelmshaven.

Aspects of classification were identified as follows:

Incident class	Environmental
Category	Wear and material fatigue
Predictability	Unpredictable
Commencement	Immediate begin
Duration of an incident	Medium Period
Measure of (economic) damage	Low
Impact	Indirect Impact
Reduction of impacts	Partly manipulable

Table 3: Incident classification Bremerhaven gantry cranes

Earthquake in New Zealand

Incident description

In New Zealand, South Island, an earthquake with a magnitude of 7.8 occurred two minutes after midnight on 14th November 2016 local time.[31] Damages reported affected all kinds of transport infrastructure.[32] The container terminal in the port of Wellington remained closed for more than 10 months. The container bridges were damaged and needed repair and sections of the port subsided.[33] Imports and exports were delayed because of the closed container terminal, transports were redirected to other ports by road or vessel (e.g. direct shipborne trade from Auckland to Lyttelton

[27] Picture: https://www.weser-kurier.de/bremen/bremen-stadt_artikel,-Kranfuehrer-tot-geborgen-_arid,1123484.html
[28] https://worldmaritimenews.com/archives/161741/maersk-karachi-ablaze-at-bremerhaven-port/
[29] https://www.verkehrsrundschau.de/nachrichten/riss-in-stahlkonstruktion-fuehrte-zu-toedlichem-unfall-in-bremerhaven-1788840.html
[30] https://www.cn-online.de/cn-galerie/bremerhaven-stillgelegte-kraene-wieder-am-start.html
[31] https://en.wikipedia.org/wiki/2016_Kaikoura_earthquake#cite_note-37
[32] http://www.stuff.co.nz/national/86418638/north-canterbury-75-quake-roads-damaged-and-blocked-public-transport-disrupted
[33] https://www.stuff.co.nz/business/95292005/maersk-to-return-to-wellington-when-centreports-cranes-are-repaired

has doubled). Since Rail infrastructure also cut off near Kaikoura, the coastal route was the only way to transport containers to some destinations.[34]

Aspects of classification were identified as follows:

Incident class	Technical
Category	Earthquake
Predictability	Unpredictable
Commencement	Immediate begin
Duration of an incident	Short period (less than one day)
Measure of (economic) damage	High
Impact	Direct impact
Reduction of impacts	Unmanipulable

Table 4: Incident classification earthquake New Zealand

Cyber-attack caused global delays

Incident description

On June 27, 2017 multiple IT-systems from various companies worldwide (e.g. Rosneft, Maersk) suffered from ransomware NotPetya which started to encrypt data as soon as activated.[35] The ship owning company Maersk published early in the morning that day first information via Twitter. Effects of this cyber-attack were systems failure and which were then shut down deliberately for security reasons. Systems controlling email and phone lines were not operating, productive systems for vessel and other operations were disconnected from affected systems. Maersk and partner companies (e.g. K + N, INTTRA) were unable to process bookings. 17 out of 76 APM terminals were affected by systems failure: either operations lasted longer than usual or parts of terminals had to be closed. Most affected was the terminal Maasvlakte II in the Netherlands.

Due to shut down internal systems, employees switched back to internet based systems and mobile communication. June 28 the Maersk booking system was successfully restarted and later that day bookings from other systems were accepted again. June 29 most terminals were operational partly with limitations. The terminal Maasvlakte II remained closed until June 03. That day, import containers were able to be fetched, it was said full operations started on July 07. [37]The overall damage is estimated to be US$ 200–300 million[38]. From first detection of the attack until full recovery it took more than two

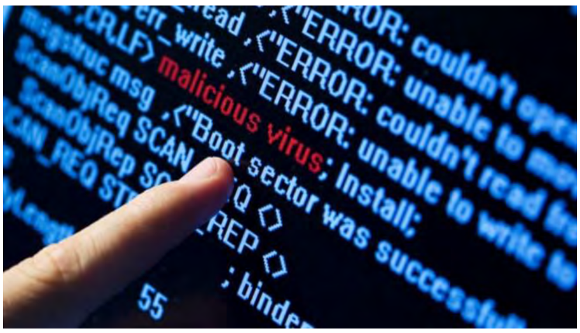


Figure 2: Example Ransom ware detection[36]

[34] https://easyfreight.co.nz/blog/nz-imports-exports-after-earthquake/
[35] https://www.golem.de/news/petya-ransomware-maersk-rosneft-und-die-ukraine-mit-ransomware-angegriffen-1706-128614.html
[36] http://www.akinsit.com/new-ransomware-notpetya-dangerous-advanced-wannacry/
[37] https://twitter.com/Maersk/
[38] https://www.porttechnology.org/news/maersk_line_reaches_profit_despite_cyberattack

weeks. The closing of Maasvlakte II created more traffic at the other Rotterdam terminals since cargo and vessels were redirected.

In a discussion at the World Economic Forum annual meeting in January 2018, the chairman of Maersk reported that 4,000 servers, 45,000 computers and about 2,500 different software programs were affected worldwide and received a software update within 10 days.[39]

Aspects of classification were identified as follows:

Incident class	Intended disruption
Category	Sabotage, Terrorism, Crime
Predictability	Unpredictable
Commencement	Immediate begin
Duration of an incident	Medium to Long Period
Measure of (economic) damage	High to fatal
Impact	Indirect Impact
Reduction of impacts	Manipulable

Table 5: Incident classification ransomware NotPetya

The history of similar incidents shows the unpredictability of such attacks. The commencement is immediate as soon as the malicious software is activated. In this case, the infected computer started to encrypt data without the ability to decrypt it again. The duration was medium to long termed since parts of the infrastructure were reactivated within minutes and other parts took weeks. The impact to the flow of goods was an indirect one since the electronic dispatch of goods was affected, not the goods nor means of transport. Such types of incidents are manipulable.

4.3. State-of-the Art techniques for event detection

4.3.1. Status information – day-to-day procedure

As a rule, status information about the respective systems is exchanged by the stakeholders involved in the supply chain during a transport. Status information provides, among other things, information about the current location of the goods or their characteristics, such as the respective customs status. The data exchange of this information differs primarily in the technical procedures. Within the framework of the ***pull method***, the information is made available by the status generator at a previously defined location, e. g. on an internal server or in the cloud, for collection by the respective authorized group of persons.

As part of the ***push method***, the information is proactively transferred from the status giver to the respective status receiver. For both procedures, automated access intervals, which can be in the range of hours or seconds, are defined between the partners. Since an exchange of status information usually takes place between different systems, such as customs systems, terminal operating systems (TOS) or freight forwarder systems along the supply chain, it is partly necessary to **convert** the respective files into a format readable by the recipient system by means of appropriate software upon receipt.

[39] https://www.golem.de/news/not-petya-maersk-itler-erneuerten-it-infrastruktur-in-zehn-tagen-1801-132407.html

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

For urgent items there is also the interrupt, or trigger method, where a status update or event may be dynamically generated and sent immediately. Also, the frequency of push update can be dynamically increased on detection of a trigger type event.

4.3.2. Container security devices and other sensors

A special type of status transmission, especially in the area of container transportation, is carried out by so-called ***Container Security Devices*** (CSD). CSDs are integrated modules, which are mounted in or on a container and serve for the recognition, transmission, logging and storage of events in the whole transport process, e. g.

- Activation, attachment and removal of the module and the sensors from the container
- Incidence of light or movement in the container
- Door opening/door lock
- Temperature, Humidity
- shock

In addition to that, CSDs are increasingly equipped with geofencing functionalities that register information such as

- Transhipment address or destination address reached/left
- Container movement inside or outside the address
- Container on the planned route or off the planned route

The data transmission to the corresponding target systems is usually carried out in real-time via GSM/GPS modules, which receive the information via the corresponding sensors in the CSD. The data transmitted in this way can be used to react immediately to events.

Figure 3: CSD ORBCOM[40]

However, the use of CSD is not yet very widespread and is primarily used for the shipping of high-value goods. Open questions such as the method of returning the CSD to the consignor, how it should be dealt with by customs, responsibility for maintenance, and the relatively short battery life are currently obvious obstacles to the widespread use of the devices.

In the refrigerated transport sector and especially in container transport (Refrigerated [Reefer] Container), such telematics sensors with monitoring function have already been established. The telematics units communicate with the permanently integrated measuring units in the container on the basis of M2M communication (Machine to Machine communication) and transmit the data via SIM card to the respective receivers. MAERSK, one of the largest container ship shipping companies in the world, had already equipped all 270,000 refrigerated *'reefer'* containers available at that time with the appropriate sensors and implemented an RCM system

[40] https://www.orbcomm.com/de/hardware/devices/gt-2300

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

(Remote Container Management) in 2017. The data monitored and transmitted within here, amongst others, is primarily current location, temperature, and atmospheric conditions inside, as well as the power status. After the data has been transmitted, it is processed internally in the RCM system and made available to the customer[41].

4.3.3. Traffic obstruction

Landside

Probably the most widespread and frequent event detection in the transport environment is related to road traffic congestion, accidents, road closures, construction sites or diversions and the associated delays in delivery which are made visible by navigation devices, radio or online providers.

The data acquisition of the different manufacturers of navigation devices or online services such as radio stations or e.g. VMZ (traffic detection panel in Germany) takes place in different ways and can be classified into two large groups.

On one hand, radio is one of the classic means of transmitting traffic data. But such messages are often inaccurate and outdated and take a long time for the user to get the corresponding traffic jams on his navigation device, as they are based on information from traffic jam detectors, surveillance cameras, fixed measurement cross sections (inductive loops, IR/radar sensors) or the police. This information may be verbally transmitted or transmitted in digital form via the Traffic Message Channel (TMC) in the inaudible part of the FM signal and processed by navigation devices.

Each traffic disruption is sent as a separate TMC message. A TMC message consists of an event code, a position, and an optional time limitation (expiration time). The message is coded according to the Alert-C standard (ISO 14819). This standard contains a list of about 1460 events such as "*lanes closed / stationary traffic*" or "*overturned vehicle(s) / traffic building up*". This list can be used to convert the message into a form that the user is able to interpret[42].

On the other hand, another method is the generation of traffic information from motion data of mobile phone users. The result is anonymous information that only reveals that someone is moving at a certain speed, but not who is moving. On the other hand, current GPS information (track data) is more accurate than information from the mobile network. Among other things, vehicles such as parcel services or taxis as well as GPS receivers in mobile phones and navigation devices with mobile radio modules provide data for this purpose. Thanks to the rapidly increasing spread of such terminal devices, there is now a sufficiently large basis of very accurate data. Data is transmitted both via TMC and increasingly - especially in high-end navigation devices - via an internal SIM card.

Figure 2 gives a compact overview of some of the possible ways in which data can be collected on the current and future situation in road traffic and shows that units which receive traffic information also provide traffic information.

[41] https://www.maersk.com/en/news/2017/06/26/maersk-line-launches-remote-container-management-for-customers
[42] https://de.wikipedia.org/wiki/Traffic_Message_Channel

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

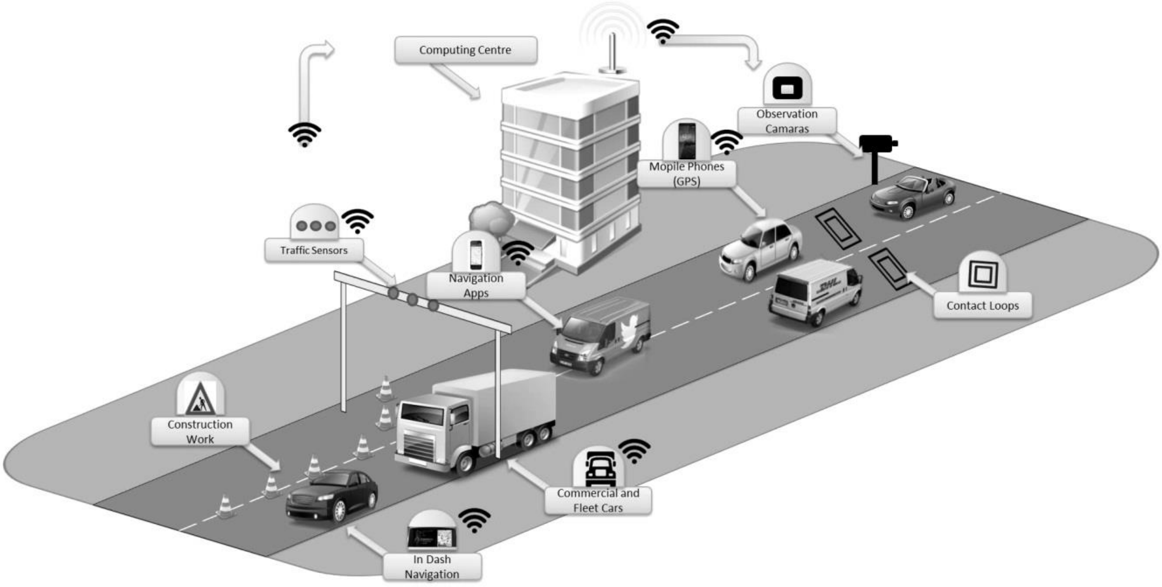


Figure 4: Traffic communication (example)

But no matter which of the different methods is used - the correct interpretation of the multitude of data by the respective provider is important. Not every longer lasting red phase at the traffic light means a traffic jam at the same time.

Seaside

In order to ensure a smooth supply and flow control of the goods in the seaports, the estimated arrival times of the vessels are determined as important events in the maritime transport environment - and partly also in inland waterway transport - using the AIS (Automatic Identification System), among other things. The vessel transmits, among other data, its current position and speed in short time intervals (2 sec. - 3 min.) predominantly on one of two available VHF marine radio channels in the HDLC data protocol for instance to the VTS (Vessel Tracking Services - Centre). The ETA (Estimated Time of Arrival) information is then calculated together with other parameters such as speed limits, adverse winds or other weather conditions. The data is decoded by appropriate software and graphically displayed. Some of this information will be provided directly to stakeholder systems such as PCS (Port Community Systems) or TOS for further detailed planning of loading and unloading processes. With the increasing availability of satellites, AIS information is increasingly transmitted via S-AIS (satellite AIS) even without terrestrial resources such as terrestrial antennas. However, the combination and logical interpretation of the different types of AIS transmission in order to create

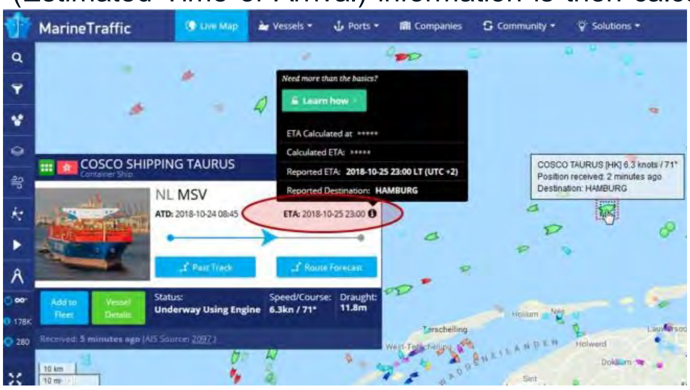


Figure 5: ETA (Estimated Time of Arrival)[43]

[43] https://www.marinetraffic.com

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

more complete and accurate information are the business environment of different information providers.

Projects such as "VESTWIND[44]" are working on making the ETA information more accurate by adding additional information such as tidal conditions, weather forecasts and traffic density analyses. The goal is a reliable prediction time up to 72 hours before ship arrival at the corresponding Port of Destination (Port of Call).

4.3.4. Environmental event detection

Environmental conditions may threaten goods transport in many ways. Some of them only cause minor interruptions of which the duration is short-lived and has no significant impact. However, other environmental conditions cause considerable disruption which, for example, partially destroys an entire port and the port environment and thus makes the transport of goods impossible in the long term.

Two important factors play a role in considering this category. On the one hand, it is the predictability that is partially given in a heavy storm, but not in an earthquake. On the other hand, the geographical position of route of transport plays a significant role. In the southern Mediterranean Sea for instance, the occurrence of a hurricane is less likely than in northern Germany. In contrast, the probability of earthquakes in the area of Cyprus is higher than in Hamburg.

Maritime and inland waterway transportation are particularly impacted by bad weather conditions like storms. Not only are vessels affected during stormy periods, but port activities like container loading and unloading can also be interrupted. Container gantry cranes, which are higher than 80 meters, and 16 meter tall van carriers are unable to function under such weather conditions. For instance, in the container terminal of the Port of Bremerhaven such activities stop at wind force of more than 8 on the Beaufort wind scale (bft).

As an example of the prevalence of this interruption, the German Weather Service (DWD) recorded 16-20 days with wind forces above 8 bft between 2011-2014, in 2015 this increased to 37 days [45]. Based on this evidence and the changing of the climate, it is logical that obstacles in transport caused by storms in the port environment will increase in the future.

From a technical point of view, meteorological stations in the different countries and regions (private as well as governmental) are measuring environmental parameters like wind, waves, current and tidal information, temperature, air pressure and rain volumes. Furthermore “Hot Spots”, which are private meteorological stations in the relevant environment, are delivering actual forecast and historical weather information. This information can be enriched by radar and satellite data to create a forecast for different spaces of time and published by relevant providers e.g. via internet platforms. In addition to that, some, like *“meteoblue”*[46], provide an API (Application-Programming-Interface) which means that relevant stakeholder systems are able to connect in a simplified way and directly with the providers systems so that relevant information can be imported. This service has a cost but the information is sometimes more accurate than information which is free of charge.

---

[44] https://www.real-eta.com/
[45] https://www.abendblatt.de/hamburg/article207740931/Der-Hamburger-Hafen-ruestet-sich-gegen-Unwetter.html
[46] https://content.meteoblue.com/en/what-we-offer/meteoblue-weather-api?

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

4.4. Conclusion “Event detection”

Apart from status information, data from events that can occur in everyday goods transport along the supply chain are predominantly collected by systems which are not at all, or only indirectly, involved in the supply chain. This data can be transferred, processed and interpreted in different ways, e. g. via an API or direct technical connections, into relevant systems such as PCS, TMS, TOS or FMS. Predictive planning or corrective measures are then predominantly carried out manually - via the human-machine interface - as the majority of the systems mentioned, especially with regard to changes, are not able to make automated rational decisions. DHL has a global ‘risk’ dashboard which tries to keep track of things like strikes or political events in addition to weather and normal interruptions.

Events which do not happen on a daily basis, but nevertheless affect directly or indirectly an unhindered transport process, such as strikes or politically based incidents are recognized and communicated in a variety of ways. In contrast to the everyday events mentioned above, they are usually not transferred to the different relevant transport systems.

5. Broad data in transport and logistics

IoT, event detection and big/broad data are inextricably linked. IoT and event detection enable vast amounts of information to be collected and analysed computationally to reveal patterns, trends, and associations through data mining and analysis, to provide insights into how supply chain operations can be improved.

The term "big data" tends to refer to the use of predictive analytics, user behaviour analytics, or certain other advanced data analytics methods that extract value from data. Traditional data processing applications are inadequate to cope with big data. The challenges include capturing data, data storage, data analysis, search, sharing, transmission, visualisation, querying, updating, information privacy, and data sources. Big data and predictive analytics are an emerging field that uses a variety of statistical techniques and computer algorithms, such as artificial intelligence, to capture industry trends and business information from large databases[47].

However, there is an extended definition referred to as “broad data” which is the enrichment of existing data by connecting to additional, new, often external, datasets. It is a value-adding technique in the area of "big data". Typically, this new data will not have been used in conjunction with any big data within a company, probably as a result of being unaware of its value. There is likely to be a connection between the parameters in these new datasets and the company big data, which means the combined data will add significant value to the business. So in broad data there is greater emphasis on original and obscure data, often external to the enterprise, to augment existing data. Broad data provides more attributes to describe existing data, enriching, increasing detail and accuracy. Consequently, the term “big data” can be misleading but it is used more commonly.

For broad data, it is necessary to look externally for original, obscure sources of data, such as those from publicly available sources and from social media. Data from regional, national and European

[47] Jeble, S., Dubey, R., Childe, S.J., Papadopoulos, T., Roubaud, D. and Prakash, A. (2017) ‘Impact of big data & predictive analytics capability on supply chain sustainability’, *International Journal of Logistics Management*

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

government agencies provide a rich source of statistics about freight movements. There are also commercially available datasets such as financial market data, weather or traffic data. Companies are beginning to understand that data has a value and that they can sell their internal data sets to other companies. The term “data is the new oil” has been coined. Examples of this might be retailers loyalty card data, telecoms location data or call patterns, companies own sales by region or market segment. All of these could be extremely useful to other companies.

Even within companies there may be obscure datasets that are little used but could help identify customer behaviour. This could include emails to and from intermediaries or customers, or it might be patterns of customer orders over time. Many enterprises are not even able to identify which customers have bought which products through which sales channels yet, so there is often some foundational work still to do with internal data.

Moreover, there can be large numbers of different stakeholder groups involved in a supply chain so there is a complexity in both the exchange of information and the diversity of data. The extent to which data is exchanged bilaterally often depends on whether a stakeholder or stakeholder group is directly/indirectly or passively/actively involved in the supply chain. The CCME (Central Command for Maritime Emergencies Germany), for example, only receives notifications regarding the transport of dangerous goods by sea - there is no feedback. Customs, on the other hand, receive customs declarations and then issue customs clearances. Each stakeholder group has different ICT systems that plan, control and support operational transport and exchange data with other partners in a variety of ways. Some internal data will be structured, and exist in well-ordered databases, but much will be unstructured, such as scanned documents or emails, or telephone transcripts. Increasingly, technology can incorporate these new sources of data in analyses.

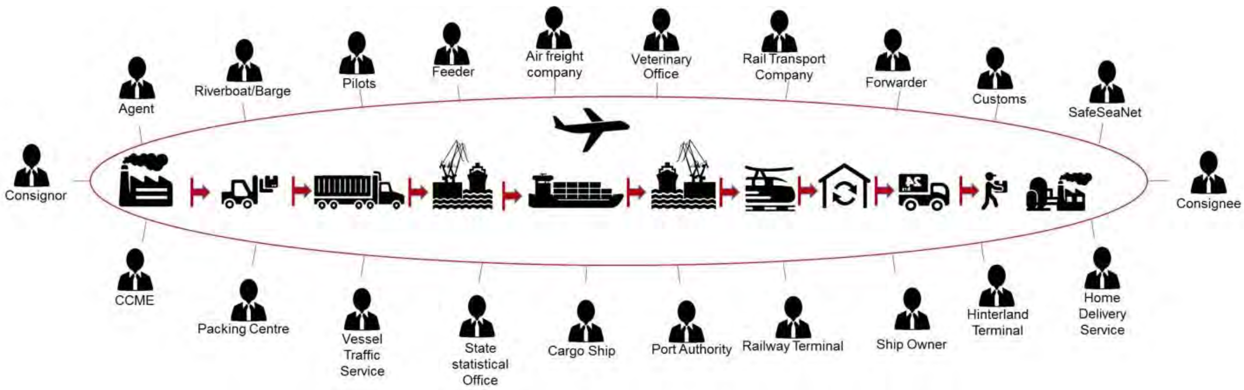


Figure 6: Stakeholder groups (fragmentary)

Artificial intelligence and big data are also inextricably linked. Big data analyses of consumer transaction data can achieve a deeper understanding of consumer spending habits[48]. In the logistics industry, by analysing data collected from customer purchasing and other activities, companies can optimise inventory while reducing delivery time and meeting consumer service demands[49]. However, many databases can contain unusable or inaccurate data, so systems need to filter the underlying

[48] Ittmann, H.W. (2015) ‘The impact of big data and business analytics on supply chain management’, *Journal of Transport and Supply Chain Management*, 9(1), pp. 1–10.
[49] Sustrova, T. (2016) ‘A Suitable Artificial Intelligence Model for Inventory Level Optimisation’, *Trends Economics and Management*, 10(25), p. 48.

data. Artificial intelligence technology combined with big data has the capability to serve the futuristic need of a fully automated transportation system with real-time solutions.

5.1. The application of broad data in transport and logistics

Big data analysis has a particularly important significance in the logistics and supply chain management of e-commerce[50]. Amazon has developed an analytics algorithm to analyse click data for tens of millions of customers browsing and purchasing goods which helps Amazon to understand customer's shopping preferences while tracking product sales[48]. With this information, Amazon can manage the inventory at each distribution centre, and reduce inventory cost while increasing the speed of delivery. Amazon also uses data analysis to determine the best procurement strategy and manage the entire supply chain process from manufacturer to client[40]. They use anticipatory shipping of products to customers before an order is placed. Data analysis helps Amazon determine when to perform joint replenishment, coordinated replenishment, and single sourcing.

Broad data analyses can support dynamic, real-time route optimisation through the intelligent correlation of data streams (shipment information, weather, traffic, etc.) enabling more efficient scheduling of assignments, optimisation of load sequences, and up to the minute prediction of the estimated time of arrival (ETA). The UPS Group has been using route data optimisation systems called Orion for many years to optimise transport routes. The software analyses fuel efficiency, driver activity, and vehicle distribution lanes to find the best delivery route and has successfully reduced distribution costs[51].

In order to be able to pre-process data for machine learning it is important to be aware which AI algorithms and for what purposes will be applied in the project. Before going into depth, considering that machine learning is interdisciplinary field that relies on many topics such as artificial intelligence, optimisation theory, information theory, statistics, cognitive science, optimal control, and many other disciplines of science, engineering, and mathematics[52][53][54][55][56]; machine learning is explained in more details.

Machine learning can be divided into supervised learning, unsupervised learning, and reinforcement learning. Supervised learning requires training with labelled data which has inputs and desired outputs. In contrast with the supervised learning, unsupervised learning does not require labeled training data and the environment only provides inputs without desired targets. Reinforcement learning enables learning from feedback received through interactions with an external environment. Based on these three essential learning paradigms, a lot of theory mechanisms and application services have been proposed for dealing with data tasks. For example, Google applies machine learning algorithms to massive chunks of messy data obtained from the Internet for Google's translator, Google's street view, Android's voice recognition, and image search engine.

---

[50] Wu, P.J. and Lin, K.C. (2018) 'Unstructured big data analytics for retrieving e-commerce logistics knowledge', *Telematics and Informatics*, 35(1) Elsevier, pp. 237–244.
[51] Sanders, N.R. (2016) 'How to Use Big Data to Drive Your Supply Chain', *California Management Review*, 58(3), pp. 26–48.
[52] Junfei Qiu, Qihui Wu, Guoru Ding, Yuhua Xu and Shuo Feng, A survey of machine learning for big data processing, EURASIP Jpurnal on Advances in Signal Processing, (2016): 67
[53] TM Mitchell, Machine learning (McGraw-Hill, New York, 1997)
[54] S Russell, P Norvig, Artificial intelligence: a modern approach (Prentice-Hall, Englewood Cliffs, 1995)
[55] V Cherkassky, FM Mulier, Learning from data: concepts, theory, and methods (John Wiley & Sons, New Jersey, 2007)
[56] TM Mitchell, The discipline of machine learning (Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2006)

Other examples include[2]:

- DHL Resilience360 – data-driven supply chain risk management
- DHL SmartTruck – route optimisation and address management using big data
- DHL Parcel Volume Prediction,
- Transmetrics – big data predictive analytics solution for transport
- DHL Trade Barometer – machine learning tool to predict air freight delays
- ClearMetal – analytics startup enabling supply chain visibility

5.2. Main advantages of broad data technology in transport and logistics

Big data can help the supply chain be more agile. A multi-agent-based supply chain management system was developed that incorporated big data analytics which exerted autonomous corrective control actions. This showed the trade-off between supply chain agility and complexity of global supply chains and how the proposed model can potentially provide enhanced responsiveness, flexibility and speed.[57]

It can also improve the coordination ability of supply chains across organisations. Gunasekaran used multiple regression analysis to prove that big data can help to improve organisational performance both in financial terms and in supply chain cooperation[58].

Big data analytics can help companies achieve customer-centric development goals. Through big data analytics, it can help enterprises to understand customer consumption behaviour, and better help enterprises optimise consumer experience[59]. Big data analytics can help the supply chain improve both visibility and transparency. Big data analysis techniques could predict fluctuation of demand in the supply chain, and combined with IoT technology, could have the potential to reduce the bullwhip effect in the supply chain[60]. Increased visibility in the supply chain can help companies have a more comprehensive understanding of their supply chain activities[61], which is a prerequisite for optimising business operations efficiency and maintenance capabilities[62].

Because professional analytics software drives big data analytics, these programs can help to reduce the complexity of data processing by linking, matching, and transforming large amounts of structured, unstructured, and semi-structured data from different databases. A comprehensive analysis of different data sources helps to understand the problem more fully and to keep abreast of development trends[63].

---

[57] Giannakis, M. and Louis, M. (2016) ‘A multi-agent based system with big data processing for enhanced supply chain agility’, *Journal of Enterprise Information Management*, 29(5), pp. 706–727.

[58] Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S.F., Childe, S.J., Hazen, B. and Akter, S. (2017) ‘Big data and predictive analytics for supply chain and organizational performance’, *Journal of Business Research*, 70 Elsevier Inc., pp. 308–317.

[59] Sreedhar, S. (2017) ‘Four Ways the Internet of Things Will Transform the Supply Chain’, *Material Handling & Logistics*, Cleveland: Informa

[60] Hofmann, E. (2017) ‘Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect’, *International Journal of Production Research*, 55(17) Taylor & Francis, pp. 5108–5126.

[61] Brinch, M. (2018) ‘Understanding the value of big data in supply chain management and its business processes: Towards a conceptual framework’, *International Journal of Operations & Production Management*

[62] Kache, F. and Seuring, S. (2017) ‘Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management’, *International Journal of Operations & Production Management*, 37(1), pp. 10–36.

[63] Julien, D., Ali, I. and Bourlakis, M. (2018) *Sedex Technology Trends Review Summary Report*.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

5.3. Challenges of implementing broad data in transport and logistics

The sharing of data is a challenge in the field of big data. According to Lambert's types of collaboration[64], in the first stage of the knowledge level, the companies would focus on short-term cooperation, such as joint distribution or line haul, back loading, etc. In the second level of the relationship, companies would be willing to share more operational information. When information is shared between companies, more accurate and efficient decisions can be made because of the ability to provide big data analytics to the breadth of information available.

Big data analysis requires a complex and advanced processing architecture and capabilities[65]. It requires people with the necessary skills and a hardware infrastructure to support advanced information processing[62]. The design of the IT infrastructure is a key issue to consider as it provides a framework for future business practices[66]. The smooth exchange of information between different departments in the entire supply chain plays an essential role in the timeliness of data[48]. Both of these require considerable capital investment to optimise supply chain resource allocation, especially for small and medium-sized enterprises. Some, however, understand the benefits data can provide. Marks & Spencer, with Decoded, have set up a data academy and 1000 employees are expected to be trained in the first 18 months. This will give them the necessary skills for a data-driven organisation, and to enable them to become change leaders in the company.

The most challenging aspect of supply chain and big data analytics is based on governance and compliance. The development and installation of a supply chain governance structure are key to guiding and coordinating big data analytics. As the most critical aspect of big data analytics across the chain, this capability is critical to achieving a consensus on common goals and setting the overall direction for future growth. How to make companies in the supply chain comply with the rules of cooperation and let these companies cooperate under a relatively fair condition is one of the challenges facing the supply chain[61].

Although big data analysis technology is receiving increasing attention in supply chain management, the adoption process of digitalisation and big data in many companies is still slow. The implementation of big data analysis within a company requires the coordination of work between a company's departments[61]. If the internal information and data coordination of the company is not smooth, it will severely affect the process of data analysis. Big data analysis still has considerable obstacles in the implementation process. Helping supply chain managers develop data-driven supply chains is one of the challenges facing the application of big data analysis technology.

[64] Lambert, D., M. Emmelhainz and J. Gardner (1999) Building successful logistics partnerships. Journal of Business Logistics, **20**(1), 165-181.
[65] Smillie, D. (2016) *Data Driven Supply Chains Navigating the path to master the mountain*.
[66] Fuchs, C. and Otto, A. (2015), *"Value of IT in supply chain planning"*, *Journal of Enterprise Information Management*, Vol. 28 No. 1, pp. 77-92.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

6. Logical infrastructures for big data analytics - modern ETL approaches

6.1. ETL in a big data environment

In computing, extract, transform, load (ETL) is a process in database usage to prepare data for analysis, especially in data warehousing. The ETL process became a popular concept in the 1970s. Data extraction involves extracting data from similar or diverse sources, while data transformation processes data by transforming them into a proper storage format/structure for the purposes of querying and analysis; finally, data loading describes the insertion of data into the final target database such as an operational data store, a data mart, or a data warehouse. A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions.

An ETL process in a big data environment has special requirements due to the expected size of the data and/or due to the nature of the source emitting new data in a constant stream. It also must be ready to deal with sources becoming unavailable and expose a well defined point of control where additional sources can easily be accommodated. In other words it should at least respect the traditional 4 V's of big data: Volume, Velocity, Veracity and Variety[67].

This results in two important features that every component of an arbitrary ETL process must fulfill. First, it must be failsafe in the sense, that the transformation process must not end in case a source becomes unavailable and secondly that all components must be capable of handling streams or batches of data in a very short period of time.

It is presumed that there are different usage scenarios within the Logistar environment that make both ETL variants, a traditional batch like ETL process (see below: Unified Views) and a kind of ETL process capable of dealing with the 4 V's of big data, described above, a valuable contribution. The expected sources vary in their respective schedule of change. Some data sources, like train, truck or other schedules are expected to change infrequently and should end up as metadata using the RDF data model to be available as an easy accessible knowledge graph. Other sources, like GPS data retrieved from moving trucks or trains will basically emit a constant stream that requires a standing pipeline, processing the incoming data.

It should be noted that many applications dealing with big data implement the lambda architecture or one of its variants. That said, at the core of such implementations[68,69] there is often a messaging system that makes it possible for an arbitrary number of consumers to work on incoming streams of data. In the case of Apache Kafka[70], those incoming data packets are also available as stream that can be accessed by any number of processes dealing with the data. While many of those processes will constitute an analytics or AI function creating a model or making predictions on an available

[67] https://www.ibmbigdatahub.com/infographic/four-vs-big-data
[68] https://www.slideshare.net/BigData_Europe/big-data-europe-sc6-ws-3-pilot-sc6-citizen-budget-on-municipal-level-martin-kaltenboeck-semantic-web-company/9
[69] https://www.slideshare.net/BigData_Europe/big-data-europe-bde-project-overview/19
[70] https://kafka.apache.org

model, it is also possible to use such a consumer as a simple transformation component within an ETL pipeline.

There are parallel projects within the EU Horizon 2020 program that are also focussing on data warehouse schemes (such as mentioned here). It maybe that some of the technologies being employed within these projects could be re-purposed for use within the Logistar project. An example of ETL implementation include two societal challenges implemented for the H2020 project “Big Data Europe”. SC3 (Energy)[71] dealt with the ingestion of monitoring data from wind turbines and SC6 (Social Sciences)[72] dealt with budget data. Both implementations were using Apache Kafka as a central tool to distribute data among all anticipated applications. Both implemented an Apache Spark[73] job to do analysis on the incoming datastream. It should be noted that in this document we are presenting any process, be it analytical or transformational as a “Transformation” component in the sense of ETL. While SC3 used said transformation component to fulfill analytical tasks SC6 did a transformation on the incoming stream of budget data in a narrower sense in that it transformed the actual data model from being column based to RDF.

It becomes clear that we are defining an ETL process as only one piece of a larger aggregation of functions that do alter data in one or the other way and that an ETL process is only one besides many others within an arbitrary implementation of a big data application.

That said, the components of a modern ETL pipeline within the realm of big data are specialised and serve specific or more purposes. An open source variant of such a process would make use of the following components. While it must be noted that an implementation of such an ETL pipeline is explicitly not bound to these tools, such an implementation should follow what those tools represent.

1. Apache Flume for data ingestion[74]

   Apache Flume is a purpose built open source tool for data ingestion. Its inner structure represent highly customizable micro pipelines consisting of a source, a channel through which incoming data is sent and a target. All three parts can be highly customized by exposing a well defined API to meet every application’s needs. There are predefined sources[75] for filesystems, like the linux filesystem, HDFS or HTTP clients, predefined channels, like the memory channel or a filesystem channel and default targets like a HDFS[76] filesystem or also Apache Kafka. An arbitrary source can be connected to any number of targets by any number of channels.

2. Apache Kafka as a messaging tool

   Apache Kafka is a messaging tool based on the publish/subscribe paradigm. It can be deployed as a failsafe high availability cluster making sure the whole implementation of the lambda architecture stands on solid ground. A consumer of an Apache Kafka topic, a logical categorisation of message streams, can access data basically in two ways, first from the very first message that came in or from the offset of the last message consumed. That means that

[71] https://www.big-data-europe.eu/energy/
[72] https://www.big-data-europe.eu/pilot-social-sciences/
[73] http://spark.apache.org/
[74] https://flume.apache.org/
[75] https://flume.apache.org/FlumeUserGuide.html
[76] http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

in case the process consuming messages fails, it is possible to resume the analytics process at the last known message or to start over from the beginning should that be necessary

3. Apache Spark as the transformation component

   Apache Spark is a well known and well-supported analytics tool. It is capable of handling streams of data and it is also very well integrated with Apache Kafka, making it an ideal companion in such an environment

UnifiedViews is an ETL tool for processing of RDF data. Unified Views is specially tailored to work with RDF data in batch mode in a non-streaming environment. This tool provides interfaces to define, execute, monitor, schedule and share processing of RDF data tasks. The tool already provides certain predefined plugins so no heavy programming is required to use this. Users can extend custom plugins to define their own plugins.

There are four types of Data Processing Units (DPUs) in UnifiedViews

- **Extractor**: can obtain data from external sources. For instance, extracts data from relational databases.
- **Transformer**: transforms the input data to the specified format. For example, transforms tabular data to RDF.
- **Loader**: Loads the transformed and curated data to external systems.
- **Quality Assessor:** produces quality assessment reports on the data.

Each DPU declares mandatory or optional inputs and encapsulates certain business logic that process the data. The input and output data units can be of three type; RDF data unit, File data unit and Relational data unit.

Extractor DPUs can be used to extract static, dynamic and streaming data coming from relational data sources. Transformers will transform these tabular data into RDF. Data quality can be achieved using the quality assessor DPUs. After all the extraction and transformation, data can be stored in any graph store or triple store.

UnifiedView comes with a predefined set of DPUs. Pipelines can be scheduled to run on a timely basis or can be triggered by other pipelines. The Execution Monitor gives detailed information on the execution of a pipeline. All data processed is stored in separate graphs and can be reviewed for debugging. The underlying graph database is configurable. By default, a built-in memory store is used but any graph database supporting the SAIL API can be integrated. The Scheduler also includes a notification system that allows information to be sent about the outcome of scheduled data processing tasks.

Frontend and backend communicate via a relational database which stores all configured information such as pipeline setups, DPU configuration, execution states, or scheduled events. To support scalability, multiple backend instances can run on different machines, effectively executing pipelines in parallel. Every backend uses its own RDF working store for storing temporary data which is produced by the pipeline during its execution.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

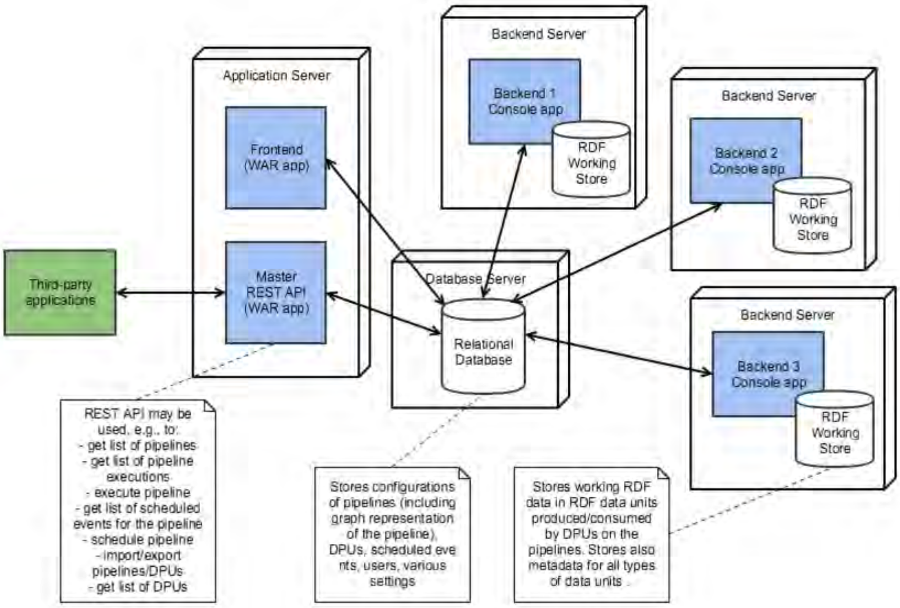


Figure 7: UnifiedViews architecture

6.2. Semantic modelling and data fusion

"The underlying success of logistics depends on the flow of data for effective management."[80]

Semantics is the study of meaning. It is concerned with the relationship between signifiers — like words, phrases, signs, and symbols — and what they stand for, their denotation.[77] Semantic modelling is a modelling technique that aims at describing entities with the focus on their meaning or interpretation in frames of the created model. For this purpose, special languages are developed and used. Semantic modelling language includes:

1. "A formal syntax and formal semantics to enable automated processing of their content."
2. "…a standardized vocabulary referring to real world semantics enabling automatic and human agents to share information and knowledge."[78]

With this language the aim is to provide a structural schema to describe the entities of interest. Such a schema is able to support standardisation, scale, and interoperability and might facilitate many tasks: search, data integration, recommender service, etc. Semantic modelling has been successfully applied in logistics[79,80,81] and related domains[79,82]. These efforts concentrate on utilizing

---

[77] https://en.wikipedia.org/wiki/Semantics
[78] Fensel, D.; Motta, E.; van Harmelen, F.; Benjamins, V.R.; Crubezy, M.; Decker, S.; Gaspari, M.; Groenboom, R.; Grosso, W.; Musen, M.A.; Plaza, E.; Schreiber, G.; Studer, R.; and Wielinga, B. 2003. The Unified Problem-Solving Method Development Language UPML. Knowledge and Information Systems (KAIS): An International Journal.
[79] Fumagalli L., Pala S., Garetti M., Negri E. (2014) Ontology-Based Modeling of Manufacturing and Logistics Systems for a New MES Architecture. In: Grabot B., Vallespir B., Gomes S., Bouras A., Kiritsis D. (eds) Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World. APMS 2014. IFIP Advances in Information and Communication Technology, vol 438. Springer, Berlin, Heidelberg
[80] L. Brock, David & Schuster, Edmund & J. Allen, Stuart & Kar, Pinaki. (2005). An Introduction to Semantic Modeling for Logistical Systems. Journal of Business Logistics. 26. 10.1002/j.2158-1592.2005.tb00207.x.
[81] L. Luncean and C. Badica, "Semantic Modeling of Information for Freight Transportation Broker," *2014 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, Timisoara, 2014, pp. 527-534.
[82] Ye, Y., Yang, D., Jiang, Z. et al. Int J Adv Manuf Technol (2008) 37: 1250. https://doi.org/10.1007/s00170-007-1052-6

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

semantics for enabling data exchange and interoperability[79][80][82] as well as enabling automatic negotiations[81].

In Brock et al (2005)[80] the authors use semantic modelling to efficiently manage the flow of data and enable data interoperability. The authors produce a system architecture that includes several schemes (ontologies) designed for several different purposes and domain experts. Yet interlinking between different ontologies guarantees the interoperability and integrity of the data. They claim that the usage of such approach “has the potential to provide unprecedented benefit and savings across industry and commerce”. The authors aim to design standards that would enable the creation of models that integrate automatically into an executing synthetic environment. The goal being to create synthetic environments that receive data from the physical world (for example through Auto-ID technology) and then produce inferences, interpretations, and predictions about the current and future states of the environment.

The combination of the proposed languages and protocols: Data Modeling Language (DML), Data Modeling Protocol (DMP), Automated Control Language (ACL), and Automated Control Protocol (ACP), represents the foundation needed to construct a general-purpose synthetic environments. The idea is that computers can construct a synthetic environment automatically, then modify it in real-time to analyze, manage and predict the states of a physical system.

In Yang et al (2008)[82] authors use semantic modelling for a similar purpose of guaranteeing interoperability of data and facilitating the data fusion. However, in addition to the previous paper, the authors come up with particular ontologies and vocabularies to accomplish this goal. Development of such controlled vocabularies as well as their manifestation as a standard and recommendation is an important activity for the semantic modelling that guarantees a wider adoption of the developed vocabularies and universal data / service interoperability.

In Fumagli et al (2014)[79] and Luncean et al (2014)[81] authors use semantic modelling to enable automatic execution of certain operations. In Fumagli et al (2014)[79] authors enable automatic execution of complex manufacturing operations. Namely, different elements of ontology correspond to different manufacturing actions so that manufacturing workflows can operate directly on the semantic models and translate those models into instructions to be executed.

The platform architecture is represented in Figure 8 by a view of its kernel module, controlling the physical system and connecting it to the upper applications. The kernel is made by 3 main layers (Physical Layer, Representation Layer and Service Orchestration Layer), plus an Interface Layer to the applications part.

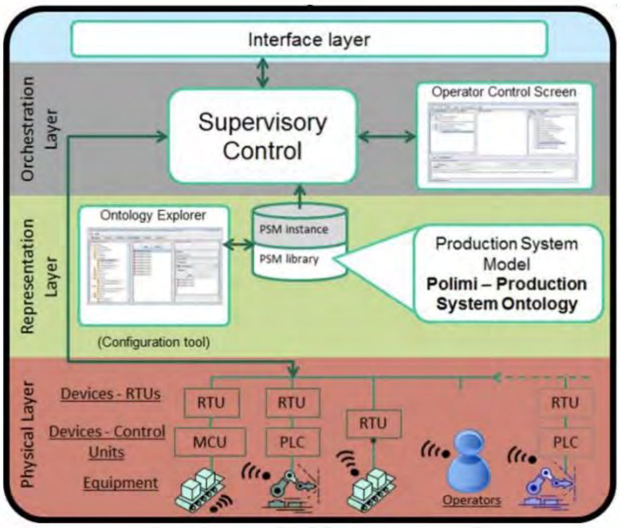


Figure 8: System kernel

In Luncean et al (2014)[81] authors use controlled vocabularies to semantically annotate services. Automatic agents can later consume these annotations and enable the process of automatic negotiations. Hence, authors describe "agent-based semantic logistics services".

Figure 9 provides the existing logistics process ontologies, their concepts, and relations.

The authors claim that the use of ontologies can be beneficial for system actors interoperability in the logistics domain, to improve communication and foster knowledge reuse, to facilitate the integration of existing systems, and to support the development process of software solutions. Therefore, they expand the existing system with four ontologies: Messages, Transport Request, Transport Resource, and Freight. In Figure 10, an example of Transport Resource ontology is provided where the concept Vehicle is expanded by its characteristics.

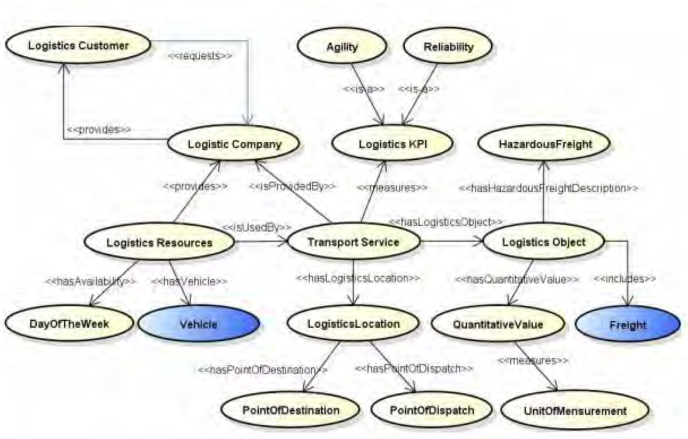


Figure 9: Logistics ontologies overview

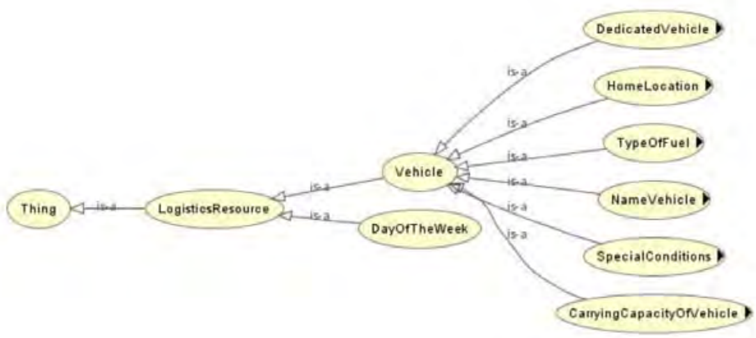


Figure 10: Transport resource ontology diagram

7. Geographic information systems

The purpose of a geographic information system (GIS) is to capture, manipulate and analyse spatial data, in particular locations, in the form of coordinates. Spatial data can be textual or tabular but must be referenced to a geographic feature. It is of significant importance in the area of transport and logistics in which GIS can be used to visually represent, for example, the locations of customers, suppliers/factories and distribution centres and their associated volumes or sizes in density maps, or real-time location of trucks through IoT sensors. One of the most ubiquitous GIS applications is satellite navigation to identify optimum routes and informing users about the current state of roads through intelligent transport systems which can improve the accuracy of travel times. It can be used to solve spatial problems, plan future operations and aid decision making processes.

GIS is made up of several components, these being the data, with software and hardware to handle the data. The most significant of these is the spatial data which can be derived from a number of sources. These could be classified into company information which may be commercially confidential and public data which may be available as open source, free of charge or purchasable. Examples of public data include current and historic data from government, mapping agencies, or other research organisations, and could include weather conditions, road delays, closures and congestion levels by time of day and day of the week, and economic conditions. Real-time current data is currently derived from the global positioning system (GPS) for identifying accurate locations on the earth's surface from satellite images.

GIS systems to handle this data can be stand-alone desktop applications or cloud-based WebMap server applications with a portal through which data can be managed and analysed. The following discussion is based around the latter GeoPortal platform which typically consists of server, client and database management applications.

7.1. Data Design

A system database and GeoDatabase architecture are essential for managing and storing all relevant available geospatial information and related contents. Equally important are data conversion and harmonisation of existing databases and datasets in order to have the same defined structure for all resources.

7.2. GeoDatabase

Geospatial database is a Geodatabase embedded as a schema within the Geoportal system database.

GeoDatabase, or spatial database, is a database optimized to store and query spatial information related to objects in space such as route plan information, statistical information, etc. While typical databases work with various alphanumeric types of data, GeoDatabases have additional functionalities in order to process spatial objects.

A Geospatial database encapsulates spatial information such as points, lines, polygons as well as 3D objects or linear networks, which are more complex structures.

Major commercial and open-source RDBMS provide spatial support (Oracle, SqlServer, MySql, PostgreSQL, etc.) with the advantage that spatial access methods use indexed structures to accelerate the retrieval and the management of spatial objects.

These extensions provide both alternative spatial representation:

- Object-based models (vector), which describes the spatial extent of relevant objects with a set of points and uses points, lines and surfaces for describing spatiality
- Field-based models (raster), where each point in space is associated to one or several attributes, defined as continuous functions (e.g. altitude, temperature, pollution, ecc.)

Spatial Data can be organized as an overlapping of layers (Image base, boundaries, roads, elevation, hydrography, land use, etc).

The huge amount of data managed and real-time data problems associated to spatial contents require methods for structuring, representing and managing data with great efficiency. This is one of the reasons why NoSQL databases, such as MongoDB, are now frequently suggested as a solution in implementing GeoSpatial Systems.

An architecture designed with mixed solution which uses the specific characteristics of both SQL and NoSQL databases, could be very effective and successful.

7.3. Application database

This database schema stores all information that is not spatial data, such as media files, documents photos, videos, etc.

7.4. Multi-Tier Architecture

The structure proposed for the GeoPortal starts from the assumption that all the inputs can be allocated into a so-called *Multi-Tier Architecture*, which assumes the subdivision of the global system into different tiers, or layers, each describing a part of the global architecture referred to as Data Storage, Business Logic and Client Presentation.

Multi-tier architecture is a concept in which the functional process logic, data access, computer data storage, user interfaces are developed and maintained as independent modules on separate platforms, allowing any of the tiers to be upgraded or replaced independently, giving scalability and maintainability to the application.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

These modules are intended to interact with each other along the lines of the client-server paradigm: the interface is the client of the business logic, which is the client management module for the presentation data. The Business Logic tier manages the exchange of information between the data source and the user interface and data visualisation. This architecture is depicted in the following figure:

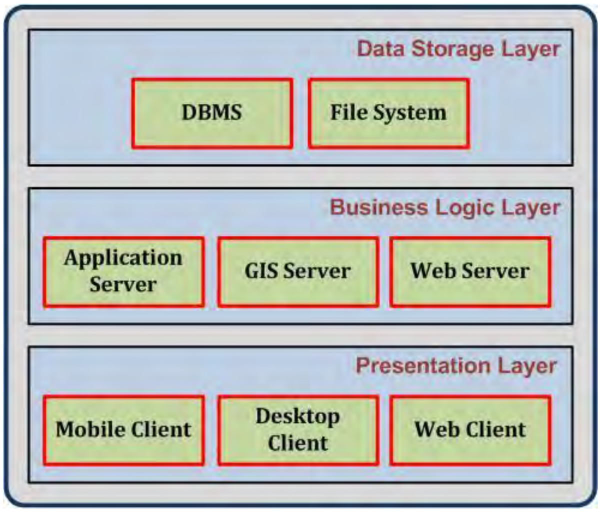


Figure 11: Multi-tier architecture

Data Storage Layer**:** This tier is designed for storing information and ensuring the correctness and consistency of them by means of a central DBMS, the GeoPortal Database.

The logical server for managing this tier is called Data Server.

- **Business Logic Layer:** This is the core processing layer, for controlling the application's functionalities and defining all the procedures that make the system work. It manages the exchange of information between data sources and user interface. The Business Logic tier hosts both Web Application Server and GIS Server, and all server-side applications. The Web Application Server permits the publication of the application and execution of the functionalities while the GIS Server allows the publication of GIS data and services.
- **Client Layer:** This is composed of the client-side applications for desktop, web and mobile devices, communicating with users sending results through a Graphical User Interface (GUI). They can be equipped with standard tools for data visualisation, management, analysis, printing, advanced functionalities such as networking, surface and/or specific applications like content management or metadata editing.

7.5. Web\Application Server

The Web and Application Server is the engine that manages the publication of general services published by GeoPortal as applications. It is dedicated to the efficient execution of core procedures (programs, routines, scripts) and the construction of related dynamic pages.

7.6. Presentation Layer

This tier is responsible for the exposure of resources to users, with the main function of translating tasks and results to something the user can easily understand.

As the Business Logic Layer includes server side applications, so the Client Layer is the location for the client counterpart. There are three possible kinds of client applications: Web application, Mobile application and Desktop application. All of them are composed by a GUI with GIS standard functionalities for data visualisation, management, analysis, printing, equipped with a series of customized menus and tools for executing specific functionalities.

7.7. Web\Mobile client applications

They allow users to interact with the system by means of a web browser, providing tools for data and map visualisation, spatial analysis, searches, queries, or printing jobs.

They do not need any client installations for they are developed as part of web browsers, whose implementation allows client-side scripts to interact with the user, communicate with server side application and eventually alter the document content that is displayed.

Web client applications can be used by all users, with different grants for accessing data and functionalities depending on the group and the role.

Mobile client applications for the proposed system have the same characteristics of the above-mentioned web applications (i.e. they do not need any client installations) but with specific requirements.

The multitude of devices form factors continues to expand, with the display screen being one of the key components. Web applications for mobile devices need to adapt to varying screen sizes, resolutions, aspect ratios and orientations. Moreover, today's devices also offer many new capabilities, such as camera, accelerometer and the possibility of GPS connection, which enhance the application field to live interaction with the territory. Applications should be able to take advantage of these capabilities in a portable manner and still deliver a rich and contextual user experience across a wide range of devices.

7.8. Desktop client application

Desktop client application allows user to interact with the system more deeply, with tools and functionalities specialized in data editing, map creation and layer authoring.

They are more powerful than Web\Mobile applications for they directly access to system database but they need a local client installation.

The following image shows an example of geoportal system architecture

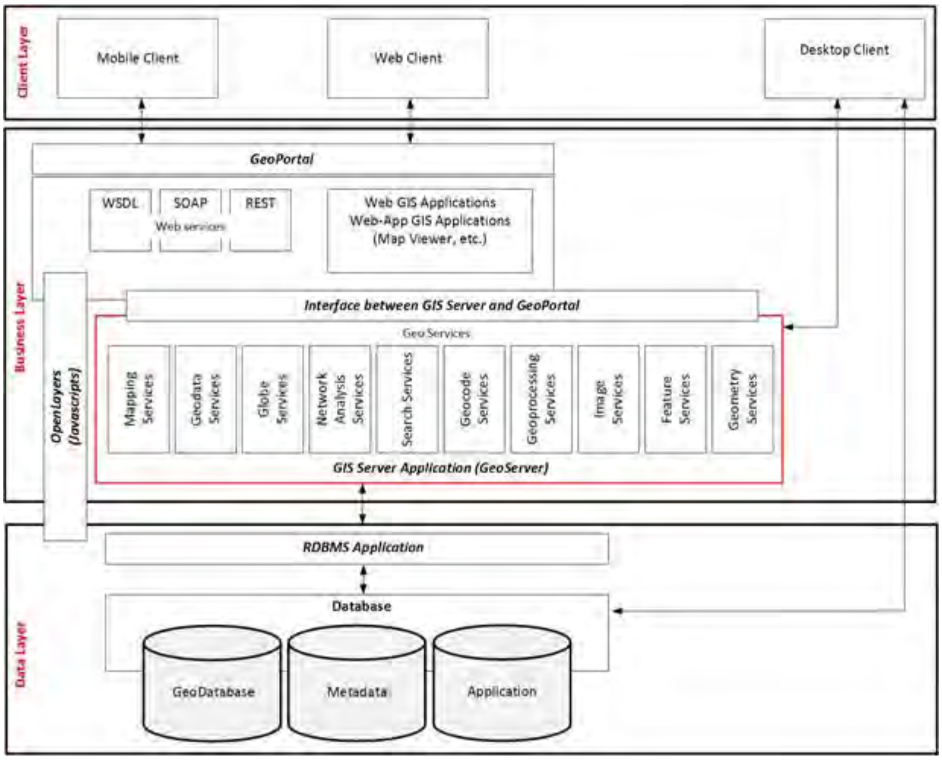


Figure 12: Typical Geoportal system

7.9. A Case Study – ROOTS Green Logistics

A COLLABORATIVE PLATFORM TO OPTIMISE RESOURCES AND MOVE THE WORKFORCE

ROOTS Green Logistics[83] is a web-based cartography system used to resolve logistics and workforce management questions, taking account of the specific conditions of each company and its customers. It plans work by optimizing travel, increasing the efficiency of available resources and balancing assigned workloads.

- **MULTI-USER & CLOUD -** ROOTS Green Logistics is a multi-user, multi-company web system for information sharing and access from any workstation, fixed or mobile. The system's cloud structure enables interaction with the systems and services of clients, telematic platforms, transport providers.

- **PLANNING -** In ROOTS Green Logistics, planning is interactive and may be automatic or manual. In both cases, the system checks that all operating conditions are met (delivery times, quantities to be delivered/picked up, capacity and typology of vehicles being used, required technical skills, driving and work schedules, breaks and stops).

- **TRACK & TRACE -** ROOTS Green Logistics monitors travel plan execution, with real-time checks on vehicle positions, delivery progress, compliance with agreed service levels. The ROOTS Green Logistics dashboard provides immediate reports for timely control of all operations.

[83] http://rootsgreenlogistics.it/

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Benefits

- It allows to control and reduce transportation cost maximizing the service level and customer satisfaction;
- It helps to reduce working time for routes planning, producing optimal route plans meeting all logistics operational constraints;
- It reduces the environmental impact of the business processes minimizing vehicles' mileage with a consequent reduction of pollutant emissions;
- Allows you to improve the workload of your resources;
- It guarantees a rapid return on your initial investment

ROOTS Green Logistics - ARCHITECTURE

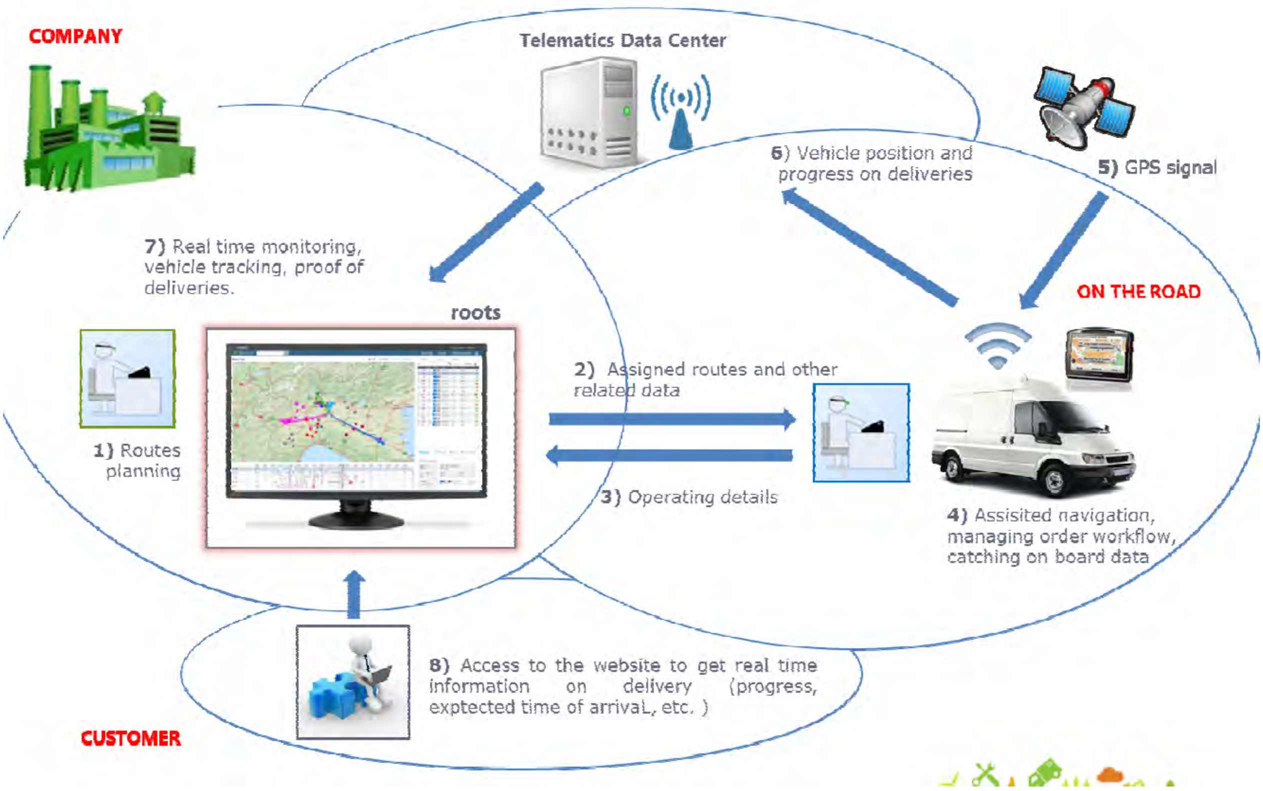


Figure 13: ROOTS Green Logistics - ARCHITECTURE

Interoperability

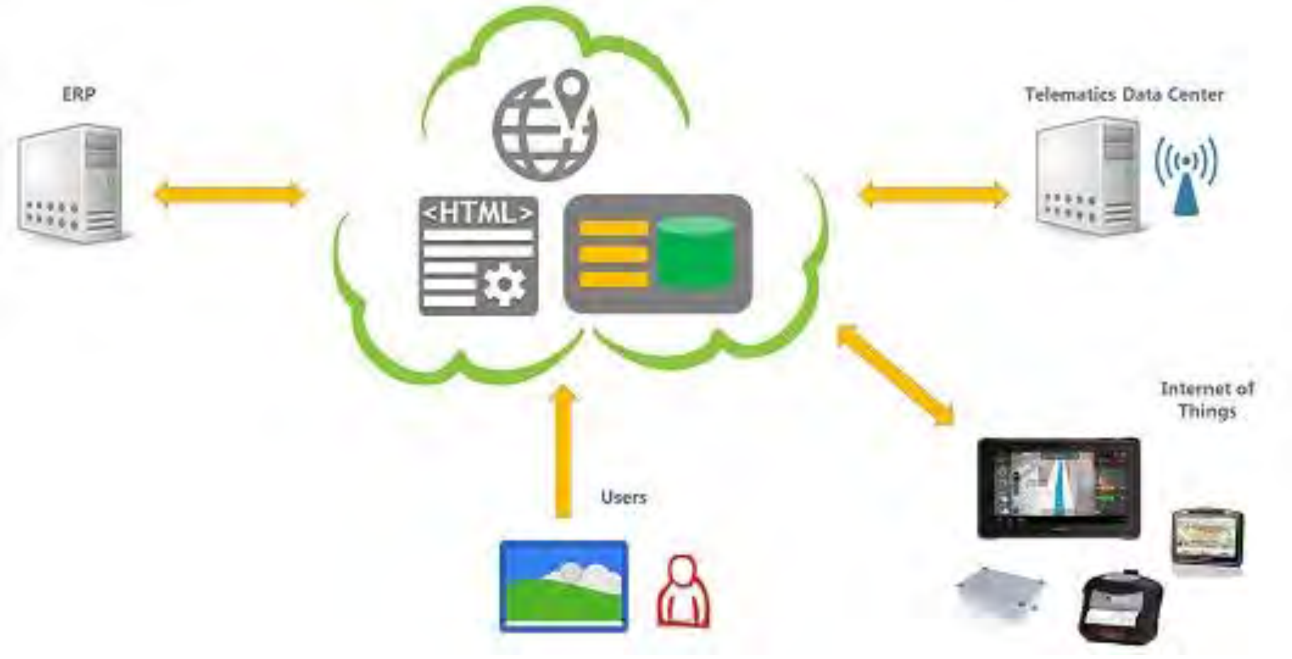


Figure 14: ROOTS Green Logistics – Interoperability

Roots is a system composed of services, web application and data transformation utilities.

It combines ease of web access and advanced integration capabilities with external services and data flows.

"FASTWAY" – a complex routing planification algorithm

- It allows users to calculate the best trip using a multi-objective function that considers both the interests of transportation (reduction of the road traveled) and of the final customer (in respect of delivery times).
- Provides a simple interface for calculating the best solution, and for evaluating alternative solutions.
- Applies to both planning and day-to-day planning scenarios to manage any unforeseen events (late departure, traffic, slowdowns in operations).
- Real-time Events/Incidents identification
- Prediction and planning re-optimisation

All monitored trucks send their positions to a data center in fixed time intervals. The positions are used for many of the services in the ROOTS Green Logistics environment but mainly for displaying the truck's latest transferred position and to track its route. The positional data is also used for data-mining purposes internally by ROOTS Green Logistics to make strategic business decisions. These are all illustrated by the figures below.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Planning Interface

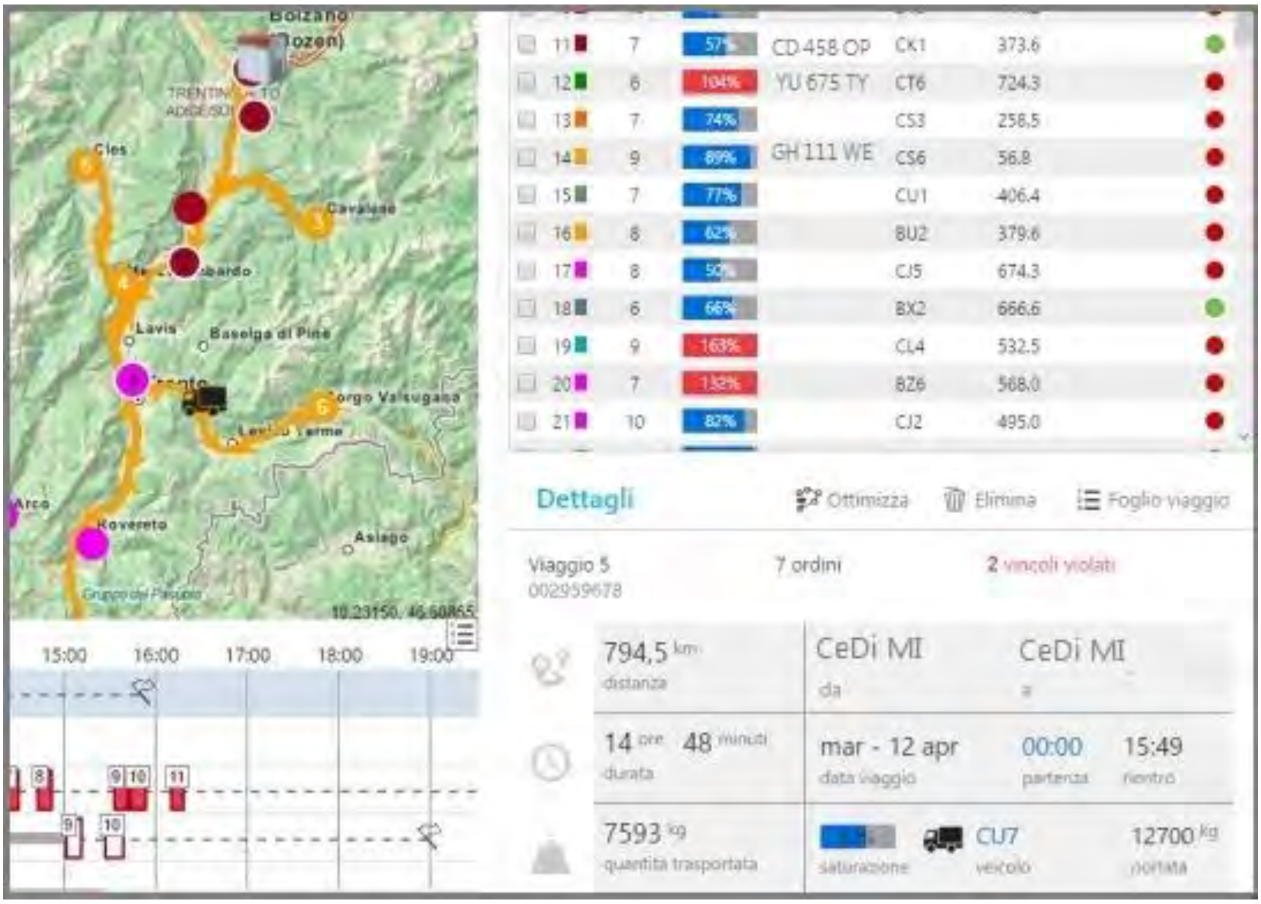


Figure 15: ROOTS Green Logistics - Planning Interface

With Roots, it is possible to have interactive planning and to plan in automatic or manual mode.

In both cases, the system will verify compliance with all operational constraints (time windows, quantities to pickup/deliver, load capacity and vehicle type, required and banned skills, driving and working time, breaks).

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Real-time Monitoring

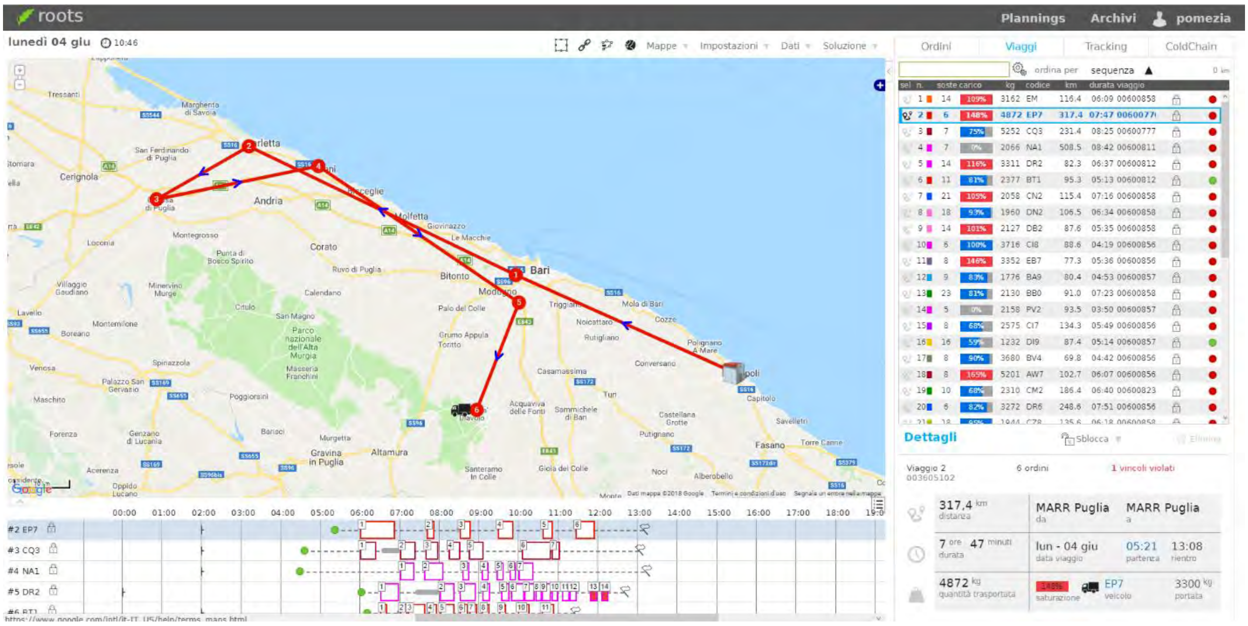


Figure 16: Real-time monitoring of the current position of the vehicle

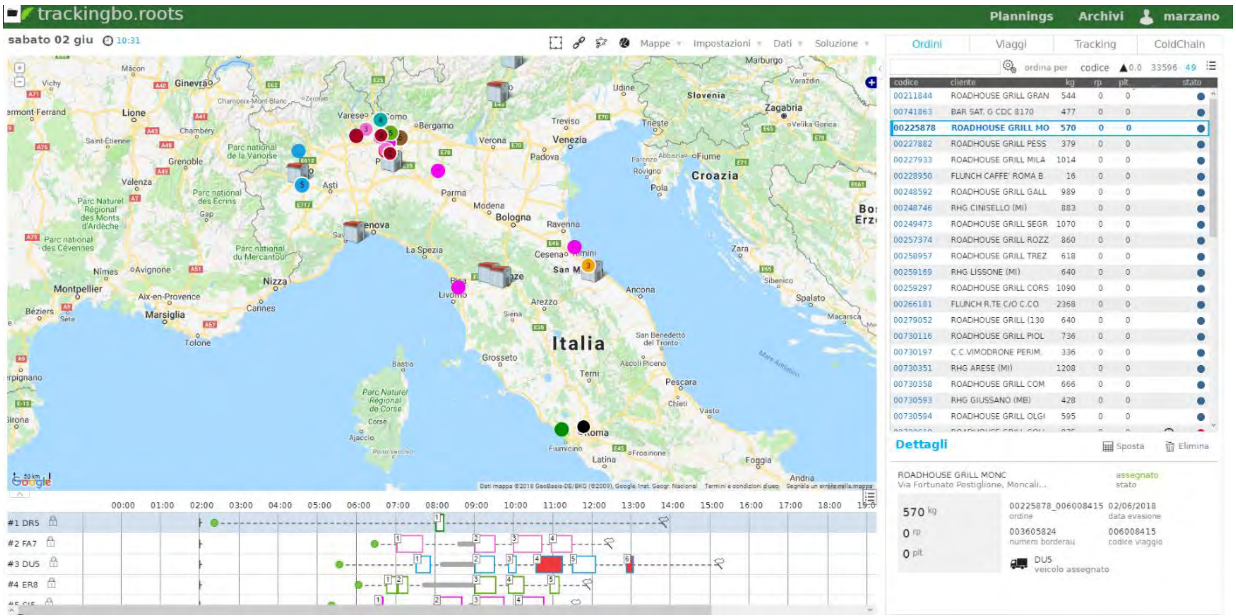


Figure 17: Monitoring of deliveries in real-time

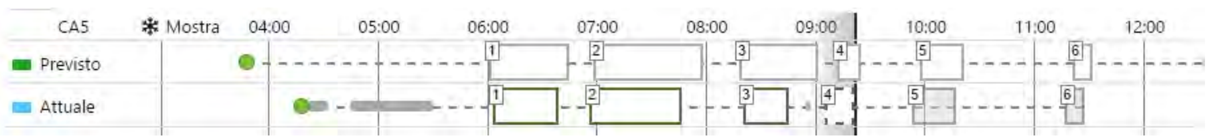


Figure 18: Estimate arrival time for customers to be served

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

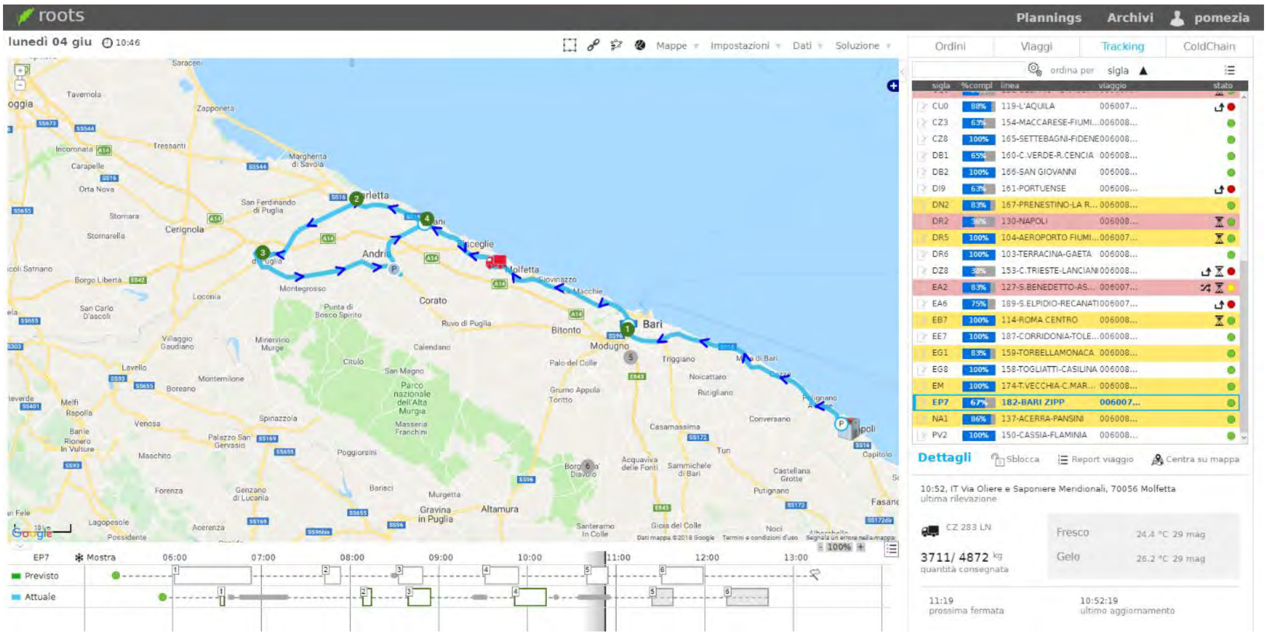


Figure 19: Real-time route recalculation and re-planning in case of logistical difficulties

Analysis of time reliability

This is measured as a comparison between the estimated delivery time possibly calculated the previous day, and the actual arrival time

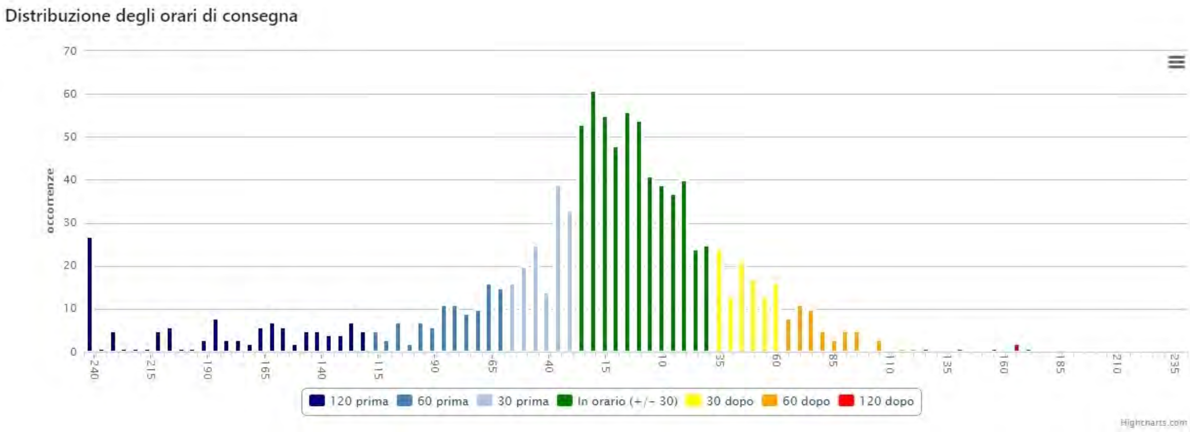
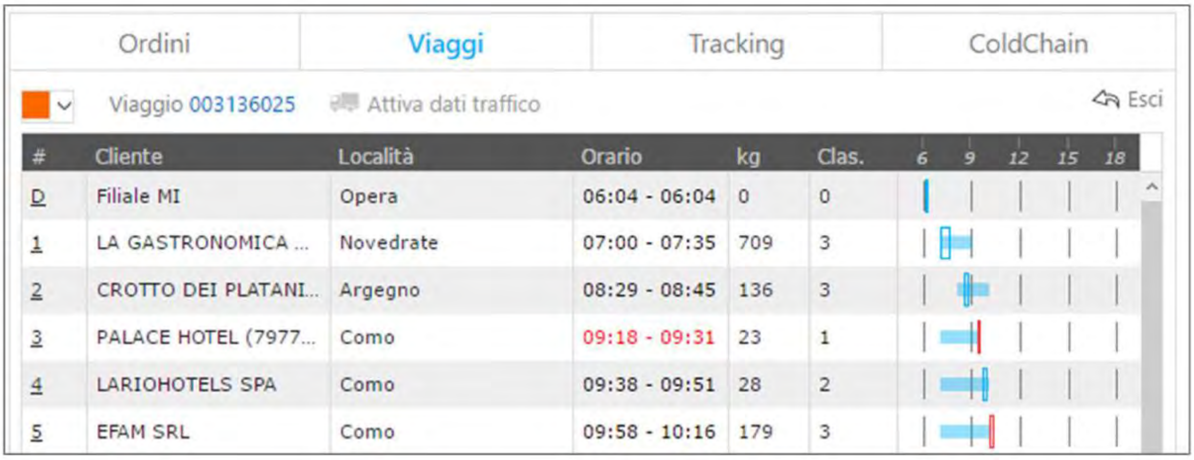


Figure 20: Analysis of time reliability

Real-time management of fleet delays

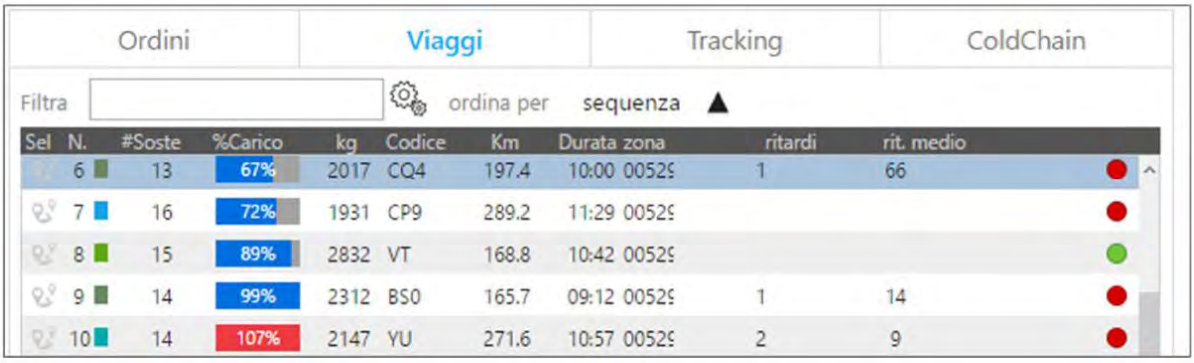


#	Cliente	Località	Orario	kg	Clas.	6 9 12 15 18
D	Filiale MI	Opera	06:04 - 06:04	0	0	
1	LA GASTRONOMICA ...	Novedrate	07:00 - 07:35	709	3	
2	CROTTO DEI PLATANI...	Argegno	08:29 - 08:45	136	3	
3	PALACE HOTEL (7977...	Como	09:18 - 09:31	23	1	
4	LARIOHOTELS SPA	Como	09:38 - 09:51	28	2	
5	EFAM SRL	Como	09:58 - 10:16	179	3	

Figure 21: Real-time management of fleet delays

As mentioned above, the truck will do the same estimation as the client to compare with the actual position. When an event occurs such as exiting a highway, or turning left in an intersection, the estimation algorithm will not only deviate from the actual position but also display an incorrect route. The algorithm then triggers a position update to prevent displaying an incorrect route in the client.

Delay details



Sel	N.	#Soste	%Carico	kg	Codice	Km	Durata zona	ritardi	rit. medio
	6	13	67%	2017	CQ4	197.4	10:00 0052	1	66
	7	16	72%	1931	CP9	289.2	11:29 0052		
	8	15	89%	2832	VT	168.8	10:42 0052		
	9	14	99%	2312	BS0	165.7	09:12 0052	1	14
	10	14	107%	2147	YU	271.6	10:57 0052	2	9

Figure 22: Delay details

One problem with estimating the position in real-time is instances when the estimation is ahead of the real truck. It happens when the truck slows down relative to the last position update and thus the estimated location will be ahead of the real truck. It is only a problem in combination with an unexpected turn or exit because the estimation will for some duration travel on an incorrect route until a new position update is done to correct this error. The duration of an incorrect route should be reduced and preferably removed entirely.

Temperature monitoring

The temperature of products in the vehicle is constantly monitored by the sensors, sent via GPRS to the TomTom server, and from there to the operations center.

Information about a vehicle's position and temperature is available in near-real-time; A tool is available for the fleet manager to enable action in a timely manner and for recording data for transport quality certification.

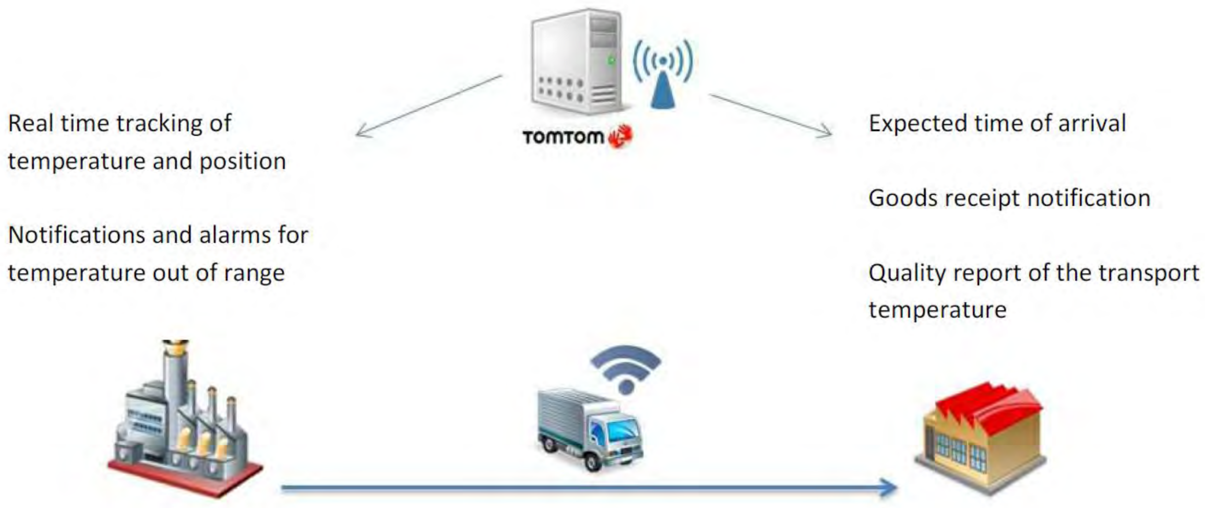


Figure 23: Real-time temperature monitoring system

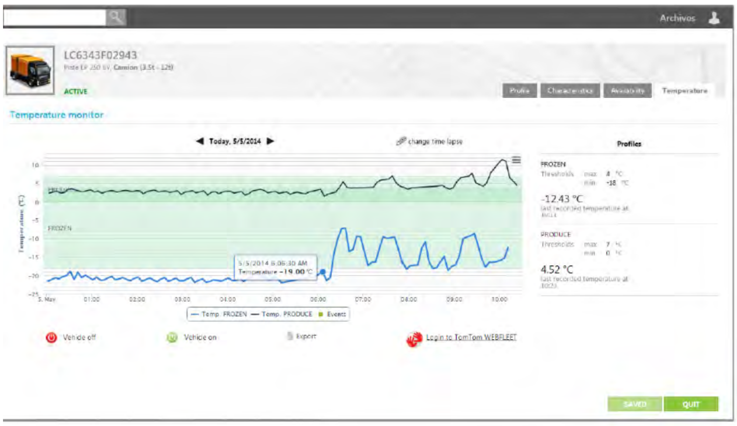


Figure 24: Real-time monitoring user interface

The user sees on the monitor the GPS track of vehicles, and graphic information of the temperature measured by the sensors. Reporting tools enable the querying of data to produce statistics on transport operations.

KPI and Reporting

Figure 25: Report example

Several indicators can evaluate the quality and efficiency of the level of service, through the comparison between planned and actual routes and there is a data summary of completed routes and deliveries.

The information collected during the execution allows:

- The comparison between planned and actual route
- The definition of indicators related to the customer service level
- The definition of indicators related to carrier service level
- These KPIs are shown in several reports and in summary emails sent to the supply chain manager

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

Multi-Users and Cloud

This is an example of a web-based application which is multi-user and multi-company.

It allows users to access and share the same data from every workstation, including mobile devices.

The cloud structure of the system enables interaction with customer's systems and services, telematics platforms and logistics service provider.

8. Conclusions

This report has examined the state of the art for a range of technologies to be used in the Logistar project. It has introduced the concepts and principles, and discussed best practices, examples and case studies. The technologies involved are cloud computing, IoT devices and systems, event detection, broad data and big data analytics, and geographical information modules.

There is increasing interest in these areas and it is inevitable that these, and other, technologies will continue to evolve with new techniques and systems being developed over the coming years.

DHL's logistics industry development trend report shows that companies' supply chains are increasingly transitioning to digital, information, automation, and intelligence. This is especially so for large technology companies such as Amazon, who are at the forefront of applying these cutting-edge technologies in the domain of logistics and warehousing management.

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.

List of abbreviations and acronyms

Abbreviation / acronym	Definition
AIS	Automatic Identification System
API	Application-Programming-Interface
CSD	Container Security Device
ETA	Estimated Time of Arrival
FMS	Fleet Management System
GeoCMS	Geospatial Content Management System
GIS	Geographic Information System
HDLC	High-Level Data Link Control (Networkprotocol)
IR	Infra Red
KPI	Key Performance Indicator
M2M	Machine to Machine
PCS	Port Community System
RCM	Remote Container Management
Roots	ROuting Optimisation and Transportation
SI	Software Implementation
SIM	Subscriber Identity Module
SOA	State of art
TMC	Traffic Message Channel
TOS	Terminal Operating System
VTS	Vessel Tracking Services

LOGISTAR project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement no. 769142.