MOBILITY MEETS BIG DATA

PROJECT HIGHLIGHTS
Introduction

TransformingTransport was one of the lighthouse projects of the EU Big Data Value Public Private Partnership which aimed to demonstrate in a realistic way the potential of Big Data for transport in Europe. Funded by the EU’s Horizon 2020 research and innovation programme, the project was coordinated by Indra, one of the leading Spanish global technology and consulting companies. The consortium represents 49 partners in industry, local government and academia from Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain and the United Kingdom. With a budget of 18.7 million euros and a project lifespan of 31 months, TransformingTransport successfully demonstrated the transformation that Big Data could bring to the mobility and logistics sector which represents 15% of the global GDP and employs over 11 million people in the EU-28 zone. Now that the project has been completed, this publication presents the key success stories that Big Data has brought to the sector.

TransformingTransport performed 13 large-scale pilots covering 7 different domains (highways, rail, ports, airports, etc), measuring the value generated by Big Data in terms of operational efficiency, customer experience and data-driven business models. The project exploited access to industrial datasets from over 160 data sources, totalling over 164 terabytes of data that provide stakeholders with additional knowledge to help them optimise their transport operations.

In other words, these pilots generated awareness on data value for the organisations involved in TransformingTransport. Stakeholders have also realised how data can provide added value to their operations and how important the data sharing approach is to improve understanding of business operations. This new way of thinking encourages open data policies that can introduce new business models and opportunities.

Many large-scale projects have acknowledged the increasingly important role that ICT plays, reflecting the major impact that the digital transformation has on the mobility and logistics sector. Nonetheless, TransformingTransport has addressed the following limitations from previous Big Data projects with respect to the use of Big Data technologies:

- Many projects address Big Data from a domain-specific single angle, focusing usually on infrastructure. This limits the potential to develop reusable and multimodal Big Data technologies. TransformingTransport demonstrated that Big Data best practices can transform data-intensive sectors such as mobility and logistics.
- In many projects, data technologies play – at most – a supportive role, facilitating communication among assets and stakeholders in the mobility and logistics networks. In TransformingTransport, Big Data technologies and the value they bring to the industrial sector took centre stage.

With all the effort that the consortium put together in this reference project, TransformingTransport demonstrated through its 13 pilots the technical and economic viability of Big Data to optimise transport network operations. The following sections highlight specific results of these pilots that show the impact of Big Data in operational efficiency, customer experience and new business models.

Based on these promising results, we expect Big Data and AI-based advanced analytics solutions will support automated decision support for operational systems. These will establish the next level of efficiency and operational improvements in the mobility and transport sectors in Europe.

I would like to thank all my colleagues from the TransformingTransport project for their contributions, inspiring discussions and fruitful partnerships.

Rodrigo Castiñeira
Indra – TransformingTransport
Project Coordinator
Three key Big Data recommendations from our Technical Coordinator

TransformingTransport demonstrated that Big Data solutions are technically and economically viable and able to transform transport processes and services. The key enabling Big Data technologies employed by TransformingTransport to bring about this transformation are predictive data analytics. Predictive analytics is a significant next step from descriptive analytics. While descriptive analytics answers the question “what happened and why?”, predictive analytics attempts to answer the question “what will happen and when?”. For example, predictive analytics may help predict whether there may be a delay in a transport process, supporting transport operators in being proactive and taking action to decrease or prevent delays. As another example, predictive maintenance takes into account predictions gained from data about the actual condition of equipment to help transport operators identify the best time to conduct maintenance tasks.

The impressive outcomes and innovations of the 13 TransformingTransport pilots are detailed in this publication. It also features three main recommendations based on the technical lessons learned regarding the use of Big Data for predictive analytics.

Data quality is a key concern for the usefulness of data analytics.

Checking and coping with missing data, data accuracy, data timeliness, different time-zones (clocks) etc. is an important but resource-and time-intensive activity. To overcome this challenge, assign a “data owner” that understands the data and the domain and is in charge of data quality and plan sufficient time and effort for data quality, data integration and refinement of data collection!

Deep learning works well without extensive hyper-parametrisation.

Provided enough good quality data is available, deep learning techniques provide high prediction accuracy without the need for extensive experimentation with hyper-parameter settings. Use deep learning to make the development and engineering of Big Data solutions more productive!

Operators benefit from information about prediction reliability.

Getting additional information about how reliable an individual prediction is helping operators decide whether to act on a prediction or not. It supports operators in finding the earliest prediction with sufficient accuracy, thereby allowing more time for proactive actions. To improve predictions, augment them with confidence intervals, error ranges, reliability estimates etc.!

Dr. Andreas Metzger
paluno, University of Duisburg-Essen - TransformingTransport Technical Coordinator

Deep learning works well without extensive hyper-parametrisation.

Provided enough good quality data is available, deep learning techniques provide high prediction accuracy without the need for extensive experimentation with hyper-parameter settings. Use deep learning to make the development and engineering of Big Data solutions more productive!
The TransformingTransport lighthouse project was one of the key Big Data related actions funded by the European Commission. As a large-scale pilot action, the TransformingTransport project had an opportunity to lead the way in showing the industry how data-driven innovation can lead to an increase in productivity, efficiency and competitiveness.

The TransformingTransport project showcased promising results in the data-intensive pilot domains representing various transport sectors, such as airports and ports. It is expected that the replicability of the innovations produced by this project would be picked up by further European sectors after the end of the project.

With regard to the large size of Europe’s transport sector, the project’s innovations have a potential to contribute to a considerable overall reduction in energy intensity of the European economy. This would have a direct positive impact on the reduction of European CO₂ emissions.

The European Digital Single Market is one of the key priorities for the European Commission. The European Digital Single Market Strategy aims to enhance Europe’s position as a world leader in digital economy. The TransformingTransport project had an integral role to play as part of this strategy to help build a European data economy via a data-driven transformation of the transport/logistics sector.

As the project is now completed, we expect that the project partners make use of the innovations generated by the project and actively aim to disseminate the results for the attention of the rest of the European transport sector. It is only in the wider adoption of these technologies and innovations that the expected impacts can be realised.

The TransformingTransport project has been also an integral part of the innovation ecosystem of the Big Data Value Public-Private Partnership (see below). Through the BDV PPP the project and its members have already played an active role in fostering the data-driven digital transformation in Europe.

An important part of the Big Data Value Public-Private Partnership

The TransformingTransport project has been a proud member of the Big Data Value Public-Private Partnership (BDV PPP).

The BDV PPP aims to create a functional Data Market and Data Economy in Europe, in order to enable Europe to play a leading role in Big Data in the global market. The Big Data Value PPP is a partnership between the European Commission and the Big Data Value Association (BDVA). The industrially led association’s objective is to ensure Europe’s leading role in the data-driven world by fostering investments along the data value chain. The BDV PPP is developing an interoperable data-driven ecosystem as a source for new businesses and innovations using Big Data, which aligns perfectly with the goals of TransformingTransport.

The BDV PPP activities address Big Data technology in terms of application research, innovation development, new data-driven business models, data ecosystem support, data skills, regulatory and IPR environments, and a number of social aspects. The value generated by Big Data technologies empowers Artificial Intelligence to link cross-cutting and vertical dimensions of value creation at the technical, business and societal levels across many different sectors.

With an initial indicative budget from the Union of 534 million euros for the period 2016-2020, the BDV PPP had mobilised 1.1 billion euros of private investments by the end of 2017. With 42 projects currently running, many new innovations are being delivered and adopted, demonstrating the impacts of Big Data in different sectors. TransformingTransport’s pilots and related emerging solutions will no doubt contribute to this drive.

Further information: www.big-data-value.eu
**Big Data: Better traffic control, less accidents on the road**

Ambitious pilot studies in Spain and Portugal have paved the way

Transforming Transport has shown outstanding potential in making highways "smarter" by using Big Data to optimise traffic and alert road authorities about unfolding road conditions and incidents.

**The Ausol Highway revolutionises how tolls operate**

The Ausol highway pilot scenario of Transforming Transport’s two highway pilots took place on the 105-kilometre Ausol Highway managed by leading Spanish transportation infrastructure company Cintra. This toll way with barriers in southern Spain is part of the AP-7 highway and European E15 Costa del Sol route. It is a highly congested semi-urban corridor that connects the cities of Málaga, Estepona and Guadiaro, running parallel to a dual carriage freeway.

Importantly, the pilot focused not only in the Ausol Highway, but also on the adjacent freeway. Within the pilot, Indra, a leading Spanish global technology and consulting company developed a dashboard which uses advanced data processing and predictive analytics to facilitate proactive decision-making for the highway’s Traffic Control Centre and optimise daily operation. This has made it possible to exploit real-time predictive Big Data analytics to forecast traffic and estimate accident probability, with the idea of preventing traffic build-up and pre-empting accidents. In a nutshell, the pilot’s triple objective was to understand mobility traffic patterns, optimise highway operation and guarantee safer roads.

Out of these three objectives, Ausol’s Traffic Control Centre focused most of all on optimising traffic operation. To this end, Ausol’s dashboard includes a traffic forecast component that facilitates daily operation based on real-time traffic analysis. It optimises daily operation by gathering and analysing in-depth data about:

1. Mobility patterns in all segments of the Highway (based on one year of descriptive models, displaying traffic and accidentality metrics every 15 minutes)
2. Traffic forecast per road segment (predictive traffic and accidentality models running in real-time for the next 15, 30, 45, 60 and 120 minutes)
3. Number of vehicles waiting at toll stations in real time
4. Toll station configuration (reversible lanes status, one-way lanes status and payment type per lane)

Based on this information, the Traffic Control Centre can relate pivotal information to toll box management and staff, including lane configuration, up to two hours in advance and based on short-time forecasts.

This technology can be very useful in real-time Demand Management, i.e. to reconfigure toll station reversible lanes and payment modes. It can add capacity when traffic forecasts predict a wave of vehicles arriving to the toll station as much as 120 minutes beforehand.
The Norte Litoral highway uses high-tech solutions to alleviate traffic

The Norte Litoral Highway runs along the northwest coast of Portugal from Oporto to Caminha near the Spanish border, with a branch heading inland, spanning between Viana do Castelo and Ponte da Lima. The highway is 109 km long, 72 km of which are refurbished and 47 km of which are new construction.

This pilot took place in close conjunction with the partners of the Ausol highway pilot in order to share the maximum amount of information and best practices learnt during that pilot. However, Norte Litoral introduces some additional challenges, since it runs under a free-flow infrastructure model, i.e. a non-stop tolling highway.

As with the Ausol highway pilot, this pilot also relies on the main data fusion techniques and machine learning algorithms (clustering, data mining, regression models, machine learning, etc.) to meet the three main business goals (understand mobility traffic patterns, optimising highway operation and guaranteeing safer roads).

As mentioned above, the highway concession is driven by service levels, meaning that the Highway Traffic Control Centre’s top priority is to keep traffic flowing on infrastructure that is being regularly serviced and is in excellent condition. To this end, and thanks to the incident likelihood feature on Norte-Litoral’s dashboard developed by Indra, the Traffic Control Centre can easily schedule all road maintenance tasks and lane closures in line with traffic and accident likelihood. This can be achieved while maintaining the highest safety levels and minimising any impact on drivers.

That said, building an accident prediction model was not possible since (fortunately) there weren’t enough accidents to train machine learning algorithms. Instead, the pilot developed an accident-likelihood descriptive model. The system uses the model to raise an alarm based on five main real-time variables (road segment, time of day, type of day, road service level/traffic flow and rainfall volume) in order to display a 5-level warning per segment for the next 15-30-45-60 and 120 minutes.

Not only do safer roads facilitate the work of the Traffic Control Centre, but they also offer more comfortable roads for drivers - an input that might be crucial when deciding which route – toll highway or the alternative freeway – better fits with the driver’s needs. In this context the pilot worked on adding value to the toll Highway by improving user perception of safety, acting on two fronts:

1. **Display real-time information:** upcoming events like slow traffic or road conditions (intense rainfall, stopped truck ahead) displayed on the variable message signalling (VMS) to control driving speed limits.
2. **Shorten reaction time to accidents.** Whether the warning comes from a real-time event or a prediction for next 60 minutes, the Traffic Control Centre adapts its response by deploying special monitoring processes in advance on a specific road segment in order to limit speed in the previous segment.

Achieving the expected degree of sensitivity for monitoring at such a deep level has led to a new technological solution that addresses the gaps and uncertainties of traditional data collection methods. The Norte Litoral pilot exploits not only existing datasets but also new data sources from a revolutionary high-tech Distributed Acoustic Sensing device developed by Indra in order to obtain reliable, affordable and updated information along the road.

While the above-mentioned technology is very promising, deeper and wider analysis is still needed to uncover its full potential. On another level, the pilot team integrated the outputs of the Distributed Acoustic Sensing device into the dashboard. This enables the Traffic Control Centre to receive incident alarms in real time and take actions to increase road safety.
Road transport contributes about one-fifth of the EU’s total carbon dioxide (CO₂) emissions. From a safety perspective, it also represents a huge challenge since every year around 26,000 people die in the EU in car accidents. The logistics sector brings with it even more challenges as customers expect reliable, punctual transport and timely delivery in their day-to-day business.

Within the TransformingTransport project and under the Sustainable Connected Vehicles domain, two different pilots dubbed Sustainable Connected Cars and Sustainable Connected Trucks developed several solutions based on Big Data analytics to cope with these challenges. They touched on areas such as air quality in urban areas, operational efficiency, logistics, and traffic safety which still suffers from high accident rates.

**Sustainable Connected Cars**

One key solution to address these issues lies in connectivity. More and more vehicles are connecting to the Cloud as new cars come off the assembly line ready for connectivity. In parallel, old cars are being retrofitted with a wide range of connectivity devices. Such new telematic data can be very valuable for analysing and processing traffic related data, especially when combined with additional data sources like road maps, weather information or brand specifications.

The Sustainable Connected Cars pilot worked on processing and aggregating part of this information – i.e. Big Data – to provide valuable services to drivers and to fleet managers through online dashboards and mobile applications. The pilot focused on reducing fuel consumption and minimising operational costs, as well as on increasing security and fuel performance by monitoring drivers’ behaviour and helping them to improve how they drive.

Interestingly, the pilot’s achievements also helped identify cars on the road that need maintenance based on car report warnings, alerts or failures sent to the Cloud. To achieve this, the pilot teams created a service that analyses these issues and assigns a severity value. In some cases, maintenance times and mileage were also considered in this evaluation. Comparisons of individual performance versus brand performance were used to detect abnormalities such as during breakdown periods.

Key features developed in the pilots include:

- Exploitation of trip-related information such as route, type of route, traffic jams, weather and driver behaviour
- Identification of 15 factors for monitoring driving performance from the fine grain (single trip) perspective to a wider perspective (weeks or months)
- Integrated API to retrieve the information from different backend systems
- Intelligent notification system that provides tailored recommendations to drivers according to their behaviour
- Assessment of faults and visualisation of traffic data
Statistical breakdown related to fuel consumption

These features, along with integration in Big Data platforms, will improve the management of vehicle fleets and encourage stakeholders such as smart highways, urban transport and insurance companies to make use of the data.

From a results perspective, this pilot led to an estimated 6% reduction in fuel consumption and in CO2 emissions when drivers followed the tailored recommendations. The estimates may vary depending on factors such as road, weight or weather conditions. Overall, the developed algorithms can help identify vehicles and trip types in order to potentially improve fuel consumption.

It also worked on better understanding planning processes of logistics stakeholders through Big Data analytics, creating visual representations of relevant data and developing relevant dashboards. To achieve this, the pilot team studied key European transport corridors and how a logistics service provider navigated these corridors, looking specifically at travel times and different handling activities such as loading or unloading at final destinations. This led to new insights that can support fleet managers and operators in improving organisation and planning, especially for large truck fleets.

The system can provide tips to drivers periodically through a mobile phone application so that they can improve their driving. If followed, the tips can help save fuel and enhance car maintenance. In parallel, the system provides a visual interface for fleet managers to control cars that produce higher emissions or present issues.

Sustainable Connected Trucks

The Sustainable Connected Trucks pilot focused on improving Trucks pilot traffic and related logistic aspects. Together with different end users such as logistics service providers and technical experts, the pilot defined its objectives and use cases, then jointly developed solutions based on ongoing exchanges among pilot participants.

Regarding time spent on the road along the defined corridors, the pilot team calculated the estimated time of arrival of truck drivers more accurately than with previous methods. Arrival times are important for planning purposes as they do not only help identify arrival to final destinations but also to interim stops at ferry terminals or other logistics hotspots. This approach helps reduce large buffer times usually included in logistics planning, leading to more efficient fleet management and therefore to cost savings.

On another front, the team also made use of satellite images – an uncommon data source for traffic analyses – to obtain interesting insights for several spots along the traffic corridors. The images helped identify vehicle types and speed, which is especially helpful for high-traffic motorway intersections, construction sites or truck destination such as airports or harbours.

These results represent an important starting point for further activities, as the pilot partners discussed expanding use cases beyond the project’s end by looking at corridors in other regions and potentially integrating the solutions into existing traffic management software already in use. Lastly, on the subject of satellite images, the team uncovered enormous potential for other fields of application such as analysing landslide activity over the long term or mapping air quality.
As train passengers, we’re on our way to important work meetings, off to visit family, having a day out, or simply nipping to the shops. When we have somewhere to be, delays can be frustrating. A key challenge for rail infrastructure managers is the ability to predict and relay failure information so that they can take immediate decisions and plan corrective actions.

Thales UK has been leading a predictive rail asset management pilot to investigate how Big Data analytics could contribute a possible solution for sharing more accurate information about the health of rail assets across a main UK rail line. Operators from each of the lines supported this project by allowing researchers access to data resources and by providing experts to help interpret the results.

In the UK Rail pilot, researchers performed Big Data analytics across three use cases, namely:

- Overhead lines (OLE)
- Point operating equipment (POE)
- Track circuits and tamping maintenance operations

Data scientists conducted investigations to validate the technical and economic viability of Big Data for predictive maintenance of these assets. Data analytics from OLE, POE and Track circuit datasets resulted in the generation of valuable decision support tools which provide cues that specific train equipment is not working properly and that immediate corrective action may be required. Results from the data analytics of Tamping Maintenance planning delivered a decision support tool which both monitors and models the degradation of the track, based on sensor data and predictive models. It then provided rail operators with information to make decisions about where to tamp based on the degradation rate of the track as a whole, the effectiveness of tamping at a location and the amount of work that a tamping machine can achieve per shift.

By implementing the respective decision support tools, maintenance activities can be targeted and prioritised effectively, which in turn will reduce the number of asset failures, reduce maintenance costs and ultimately maximise operational efficiency of the railway.

Similarly, Indra has been leading a pilot, based on a ‘high speed’ line in Málaga, Spain, also consisting of three Use Cases: point machine, track profile, and operational restrictions. The objective of this pilot was to improve the reliability of high-speed rail networks by optimising operator performance and maintenance of the rail infrastructure.

The importance of applying Big Data and deep learning technology to optimise railway infrastructure and specifically high-speed railway maintenance and management has proven to be a success, not only because of the optimisation of the railway’s condition but also as a way of exploring the numerous possibilities that employing Big Data technology offers for improving railway operations. One of the most important points of employing Big Data in high-speed railways is that new sources of information can be integrated, and the predictive model can be easily optimised to show new data, thus keeping the railway maintenance model up to date. The main advantage that this new maintenance model has compared to the current maintenance procedure is that it calculates predictions for railway maintenance and management with enough time to schedule and optimise each task needed for the correct operation of the high-speed railway while minimising passenger disruption.

The full process to attain the new predictive maintenance model starts with obtaining data from a dedicated train named Seneca, which analyses track conditions and displays predictions. To accomplish these predictions, the most important variables that affect the track conditions and safety parameters are selected for analysis by the developed algorithms. Also, past events and maintenance tasks previously undertaken are included and stored so that as many variables as possible are taken into account.
Big Data offers important opportunities for making port and terminal operations much more efficient. Currently, barely 3% of ports are automated. However, as cost-effective solutions for automation are emerging, this facilitates streamlining port operations. One particular innovation that will contribute to increasing efficiency is productivity cockpits that offer predictive data analytics and decision support capabilities. Together with novel sensing devices these cockpits will also deliver new predictive maintenance solutions that, for example, can proactively alert operators about faulty components in cranes or trucks and help prevent production downtime.

These and other innovations are part of the TransformingTransport project, which has successfully applied Big Data analytics solutions in two pilot scenarios – one pilot in Valenciaport (Spain) and the second one in duisport (Germany).

**Terminal Productivity Cockpit (TPC)**

The main idea behind both pilots was to develop a Terminal Productivity Cockpit (TPC), a web dashboard to monitor and improve KPIs such as truck turnaround time and train punctuality by revealing key trends. The TPC exploits advanced data processing and predictive analytics techniques to facilitate proactive decision-making and to streamline operations. It was designed to monitor port terminal processes and equipment to proactively trigger maintenance tasks and adapt relevant processes.

Predictive Terminal Process Analytics leverages cutting-edge predictive business process monitoring solutions, i.e. real-time predictive Big Data analytics for terminal processes, in order to follow container movements in the terminal (see heat map image). Several stakeholders were involved in mapping container management operations, providing better insights and metrics about overall efficiency at terminal level.

Predictive business process monitoring anticipates performance issues in ongoing processes, looking at how each process unfolds as it nears completion. If the Big Data solution predicts an issue, operators can proactively adapt the process to overcome it. For example, if the system predicts a delay in departure time for an outgoing train, operators can change container handling priorities or employ other transport equipment before the delay actually occurs. In this context, the TPC provides predictive analytics results to terminal operators and to dispatchers in order to enhance decision support.

Predictive Terminal Process Analytics must achieve a high accuracy rate (e.g. avoiding false negative predictions) but also generate predictions as early as possible in order to increase proactiveness. However, there is an important trade-off between these two requirements as late predictions typically have a higher accuracy. The pilot ultimately addresses this trade-off using ensembles of deep learning models (e.g. recurrent neural networks) and enhanced computational results. For example, some of the KPIs obtained imply a monthly average reduction of pollutant emissions between 15%-25%. Other KPIs which are more focused on maintenance tasks imply an average reduction of more than 15% in the monthly number of maintenance interventions.

In addition, the schedule of these maintenance tasks implies an optimisation and facilitation of the railway management due to improved railway maintenance which is directly related to railway traffic. Predictions given by the model inform the user of the location of the specific track section where maintenance is required. Moreover, the tool can inform the user if another maintenance task would take place in a section within close proximity to the first one and within a short period of time. Therefore, both sections can be simultaneously repaired, reducing maintenance time and costs.

Overall the optimisation of maintenance activities directly affects traffic management. Reducing the number of maintenance interventions will increase the number of trains that can ride daily over the railway tracks, resulting in an increase in the number of train trips and therefore more transport possibilities for rail customers.
prediction models, leading to a cost savings of 27% on average.

In parallel, offering total integration into the TPC, PEMS minimises and prevents equipment and system failures. It predicts maintenance and replacement of equipment based on facts and usage rather than on estimates. This helps mitigate delays and increases the effectiveness of terminal operations, complementing the predictive capabilities of Predictive Terminal Process Analytics and enhancing terminal productivity. The result is better support for port operators and an improvement in relevant KPIs.

Predictive maintenance in cranes

Predictive maintenance in both pilots has focused on container cranes or spreaders, and specifically on crane twistlocks, the mechanism that locks a container into the crane. Typically, the twistlocks are replaced after a certain number of pre-defined moves without taking into account actual wear and tear. More data on crane operation can certainly boost predictive maintenance, keeping in mind that twistlocks are the primary reason for crane failures.

When a crane fails due to faulty twistlocks, it hampers the logistics process. Thanks to predictive analytics based on machine learning techniques, port operators can now be alerted in advance when a crane requires attention or is about to fail.

Importantly, when it comes to monitoring twistlocks and cranes, the TPC visualisation dashboard employs easy-to-read gauges (see image), providing a quick user-friendly summary of the current status and key data. The gauges display current windspeed, temperature and power consumption separately, along with fault probability.

Progress in exploiting Big Data for ports

Both pilots improved their analytical models according to end user feedback, ensuring better accuracy of the selected KPIs. In addition, the pilots put the mechanisms in place to exploit live data and not only rely on historical data. They highlighted potential application of Big Data analytics under real port conditions, demonstrating significant potential savings in time and money.

Against this backdrop, the Valenciaport pilot applied predictive maintenance to the port’s gates by analysing what happens when gate cameras malfunction. Normally, camera failures mean that truck and container identification must be done manually, creating a domino effect that involves other trucks. This in turn leads to long queues and considerable delays, negatively affecting port operations. The pilot’s analysis confirmed that these delays or waiting times are directly related to gate downtime during peak hours. Within this context, the pilot successfully predicted gate performance based on indicators such as camera reading reliability, helping operators to take action before failure occurs. This includes alerting staff when a camera requires checking or maintenance (e.g. cleaning the camera lens or changing its angle) and redirecting traffic to other gates.
Passenger satisfaction is currently one of the most significant indicators for benchmarking airports and airlines. It provides valuable information on airport, passenger and airline processes that impact effectiveness and efficiency. Passenger satisfaction can also offer detailed analysis, trends, pattern recognition and solutions in retail revenue development and marketing based on personal preferences.

**Enhanced business models at Athens International Airport**

In the first of two aviation pilots under TransformingTransport, Athens International Airport (AIA) came together with Indra, Aegean Airlines and Airport Gurus to analyse the impact of Big Data on smart passenger flow and retail. The pilot, which took place at AIA and was led by Indra, formed a partnership that combined the expertise of airport operations, the local carrier and ground handling companies. These synergies not only helped in sharing insights on passenger operations among airports, airlines and ground handlers, but also enhanced data exchange to understand and consider customer needs for improving airport retail business.

The first phase of this pilot focused on understanding passenger flow within the airport, highlighting its impact on other airport processes such as aircraft turnaround, security screening and retail. Descriptive models that describe passenger behaviour were defined based on historical data. The combination of these models with flight schedules for a specific period enabled the pilot to predict when each type of passenger (business, economy, groups, etc.) arrives at the airport and how that passenger circulates within it. Using such Big Data analysis can help airlines and airports make better decisions about the passenger process, improve passenger flows by identifying areas of improvement and ultimately increase passenger satisfaction.

In the second phase, pilot stakeholders teamed up with the airport’s Duty Free operators to define and understand different variables impacting passenger behaviour and sales. By analysing retail sales per passenger and flight segment of a specific flight, as well as the retail spending behaviour of each flight segment per passenger, the pilot team successfully identified best performing flights and segments in terms of sales. These insights can be extended even further with retail basket data analyses to provide information on passenger retail preferences and create personalised offers that also enhance passenger satisfaction. The pilot also focused on retail sales based on flight allocation at different parking stand categories. By identifying the spending patterns of passengers per flight and assigning those flights close to the airport’s retail area, the pilot can maximise retail revenues for the retailers while increasing dividends for the airport. This represents an excellent example of how the airport can maximise retail sales.
Better aircraft turnaround at Milano Malpensa International Airport

Flight punctuality or on-time performance (OTP) is essential for improving passenger experience. This doesn’t only apply to hub airports where on-time departure of final-destination flights depends on the punctuality of incoming flights and on the smoothness of transit operations. It also applies to smaller point-to-point airports favoured by commuters. In both cases, flight punctuality is the result of a non-visible process that involves several steps in airside operations.

The second aviation pilot took place at Italy’s Malpensa International Airport with the involvement of SEA, Società per Azioni Esercizi Aeroportuali, the company managing the Milan Airport System. Under TransformingTransport, SEA partnered with Jeppesen and Boeing to test the impact of Big Data technology on aircraft turnaround processes.

The Milano pilot focused on the airside processes where, even an apparently insignificant improvement in accuracy of the timings expressed in minutes, means significant savings. Key timings that have the highest impact on the pilot are Estimated Time of Arrival (ETA), Taxi-in/out and Gate occupancy. To improve these, the pilot team experimented the new models in the shadow system that operates in parallel with the Malpensa Airport Information System (MAIS).

In the first phase of this pilot, the team selected the IAG Group and the Lufthansa Group of airlines to obtain Big Data. It populated the initial database with data on key timings from 2016 to 2018. After analysing the data, the pilot team created initial models which were integrated into the shadow system.

The emerging Big Data model immediately showed good results in achieving ETA improvements. While it did not achieve the expected results for taxi-in of arriving flights, it achieved better performance for taxi-out of departing flights compared to the static tables previously used to estimate the times. In addition, the model for boarding time/gate allocation time showed much better performance compared to the set of time estimations traditionally used.

Bringing aboard Sky Team and Air Italy in the second phase, the pilot succeeded in increasing the model’s performance to enable easier scalability of the results. This actually translated to over 40 flights for Sky Team and over 20 for the Air Italy Group, thus adding the two groups’ data-sets from 2016 to 2018 and then analysing the data.

Big Data impact and possibilities

Big Data has proven to be instrumental in the analysis of elements that affect terminal and airside operations, thanks to new models that analyse the current situation and predict future performance. These models are slated to become increasingly more refined, producing more effective solutions that highlight problems and pre-empting them before they arise.

With respect to the Athens pilot, the Big Data model has helped understand current passenger needs and gain insight on passenger flows, behaviours and segments, leading to better terminal design and optimised use of airport terminal resources. With respect to retail, the pilot’s successful passenger segmentation and clustering helped identify different passenger behaviour trends, which could lead to new personalised services that improve passenger experience and increase sales. For Malpensa, this has led to operational efficiency applied to the everyday airport reality.

While the solutions proposed are specific to Athens and Malpensa, the models can accept data from other airports. They can be replicated with the expert knowledge of the local operators to obtain a tailored solution which can be used beyond the scope of TransformingTransport.
Maintaining a balanced traffic flow in the city while keeping the city attractive for both travellers and commerce is a major challenge for city authorities. In order to ensure the viability of logistics operations for both logistics providers and their customers, vehicles delivering freight in the city centre should have easy access to parking places without obstructing traffic. Big Data solutions can assist local authorities in optimising the number and location of parking spaces and in improving situational awareness of urban traffic management centres.

The TransformingTransport project piloted solutions for freight delivery and urban traffic management in two medium-sized European cities: Tampere in Finland and Valladolid in Spain. Tampere, with a population of about 230,000, is located on the isthmus between two major lakes, representing the third largest city in Finland and the largest inland centre in the Nordic countries. Valladolid is located in the northwest of Spain and is the capital of the Castile and León region, boasting a population of about 300,000.

Improving situational awareness

In Tampere, the pilot developed tools for improving the situational awareness of urban traffic management centres. These tools rely on the different sensor data sources that contribute to analysing and modelling the traffic in the city. This includes fixed sensors such as loop sensors at traffic light intersections and permanent traffic counters on arterial roads, as well as traffic cameras installed at critical points. To this end, the city of Tampere has made significant effort in providing open data to encourage new services that can benefit from exploiting such data. Open datasets include aggregated loop sensor data and real-time traffic data from public transport vehicles. The partners in the project include VTT Technical Research Centre of Finland, Infotripla, Mattersoft, Taipale Telematics and the City of Tampere.

Infotripla developed a fluency model for traffic in the city and a dashboard to assist the urban traffic management centre in day-to-day operations. The dashboard shows the state of the traffic based on analysis of static sensors and GNSS data from vehicle fleets. The dashboard is a low-cost solution which can be replicated in smaller cities without the need for investing in a traffic management centre. VTT developed a tool for analysing traffic camera feeds based on artificial intelligence technologies. If the system identifies a traffic jam, it automatically notifies

© Visit Tampere, Suomen Ilmailukuva Oy
the traffic manager. Social media, especially tweets, are scanned for traffic-related messages from the Tampere area. Tweets containing selected keywords are shown to the operator and may contribute to decision making to help alleviate traffic.

Based on the available information, traffic managers make decisions regarding how best to mitigate traffic and inform travellers. Infotripla investigated new ways for providing the information and looked at how to automate its dissemination. For critical traffic events such as tunnel closures or police-reported accidents on the major roads, the system can send automatically computer-generated twitter messages.

**Sustainable Freight delivery**

The main objective of the Valladolid pilot was to guide city authorities in making decisions on how to improve urban logistics in specific areas of the city that see intense freight transport activity. To achieve this, it considered traffic simulation models which are very useful for traffic managers and city authorities since they help them analyse the impact of new or upcoming traffic-related regulations (e.g. regarding loading and unloading of goods). Such models are also crucial in demonstrating to politicians and citizens that the decision-making process is being supported by fact and is objective. The partners involved in the pilot are CARTIF Technology Centre, PTV Group, TomTom LTG, Grupo LINCE and the City of Valladolid.

Specifically, the pilot developed two traffic micro-simulation models (see dashboard images) and simulated different scenarios. It also developed a planning tool for delivery fleets, along with a dashboard that displays key insights. The tools were developed by CARTIF and are available online at tt-valladolidpilot.cartif.com.

In addition, PTV has used macroscopic approach to analyse emerging data sources and generate origin-destination (OD) matrices that are useful for extended traffic modelling in Valladolid’s urban and interurban areas.

In Tampere, Mattersoft developed and demonstrated an application that allows logistics providers to reserve designated parking spaces for freight delivery. The driver can reserve the parking space through a mobile application, while traffic wardens can verify if the vehicle is permitted to park at the parking space.

**In a nutshell...**

The two pilots have demonstrated the potential of Big Data for improving traffic in the city. By exploiting the data generated from the ever-increasing number of sensors, urban traffic management centres and city authorities are better aware of the traffic situation. There is no doubt that modelling city traffic can help develop alternative solutions to support city authorities.
Redefining e-commerce logistics

New Big Data services help streamline the online delivery and pick-up

The findings of the Dynamic Supply Networks pilot present various insights about the role of Big Data and business analytics in e-commerce logistics, along with their potential impact on improving logistics processes, as well as on providing innovative services to enhance customer experience. Within this context, six scenarios were formulated in association with different application domains:

- **Scenario 1**: Identifying delivery patterns and problematic issues, as well as forecasting
- **Scenarios 2 and 3**: Shared Micro-hubs and Click and Collect (C&C) points solutions over the largest Greek logistics provider network of shops/warehouses in the Athens region
- **Scenario 4**: Revisiting the replenishment and delivery processes of a brick and mortar grocery retailer under the light of online grocery by combining data sources from major Greek grocery retailers and suppliers
- **Scenario 5**: Online consumer insights regarding delivery processes deduced from consumer reviews that were extracted from various online shopping aggregators
- **Scenario 6**: Geographical segmentation based on shoppers’ ordering behaviour

By considering the requirements of each scenario, the pilots developed and deployed a series of services to support end users in decision making. This includes analysing the current distribution processes of a third-party logistics (3PL) provider or inspecting final consumer views regarding distribution processes. The scenarios exploited data from the project’s 3PL partner, top courier companies and online and brick and mortar retailers in Greece, as well as open data and various social media sources. The pilot pallet image presents various data analytics solution deployed along with the data used.

More specifically, the Big Data infrastructure developed supports data manipulation and business analytics, yielding a unified dashboard which displays key services of the pilot’s system. The services help potential end users reach different decisions which range from analysing current distribution processes to inspecting final consumer views regarding distribution processes. Some technical components and functionalities of these services are:

- **Descriptive analytics dashboard** that interacts with third-party logistics delivery data by selecting various criteria or dimensions and displaying results visually
- **Forecasting analytics dashboard** that predicts the behaviour of different types of delivery data, depending on the selection of various dimensions, and presenting forecasts through simple graphs
- **Routing algorithms for 3PLs** to optimise various objectives (e.g., capacity per vehicle, number of vehicles used, etc.)
- **Clustering-based business analytics** to exploit delivery data and identify regions with similar shoppers and delivery patterns offering strong support to managerial decisions

Applying these services led to the following:

- **New logistics network models** that overcome location allocation problems using delivery data and delivery networks of various logistics companies, integrating as well actual warehousing and transportation costs and other geographical data sources
- **Shared logistics with shared hubs and C&C micro hubs** in a metropolitan area that apply optimisation algorithms, providing an efficient sharing approach (costs distributed among selected warehouses) while achieving minimum total distance and delivery cost in warehousing and transportation. Features a dashboard with several input and output views, allowing users to create new scenarios based on different data sets and configurations, and to compare historical and new data by presenting the frequency of location decisions through heatmaps
- **Partnership between online retailer and multistore physical retailer** with locations that act as intermediate hubs to minimise transportation costs, based on a decision support tool to identify optimal hubs (ma-
chine learning approaches help to cluster locations)

- Processing large set of orders and routing data from grocery retailers and suppliers by computing various vehicle routing objectives. Datasets and algorithms are integrated into a ready-to-use system that presents suggestions according to specific parameters (e.g., the number of hubs to be used)

- Identifying critical problems in online orders and product deliveries by collecting online reviews which are classified in various problem categories (based on testing several classification algorithms)

- Geographical customer segmentation: A data mining and clustering approach that introduces a new type of geographical segmentation based on customer ordering behaviours and patterns. This approach offers behavioural segmentation and characterisation of shoppers’ orders by examining product categories included in each customer delivery. It also identifies regions with similar order behaviour. The results are then presented visually in a dashboard

Overall, the pilot proposes a framework of novel Big Data and business analytics processes and services which can greatly contribute to addressing new challenges in e-commerce logistics, while also creating the necessary deployment conditions for various logistics processes. The framework is also supported by interesting KPI results. The quantitative results together with the main technical and non-technical achievements and lessons learned also form the basis for defining the business impact and the post-project replication strategy. There is no doubt that Big Data can transform the way that e-commerce is conducted, thanks to the pioneering efforts of this pilot.

### Measuring the impact of Big Data on transportation per pilot

The evaluation framework of TransformingTransport was based on an analysis of multiple criteria to explain how Big Data technologies can help enhance the overall efficiency of the transport sector. To this aim, thirteen pilots grouped into seven different transportation modes or “domains” – Smart Highways, Connected Vehicles, Proactive Rail, Intelligent Ports, Smart Airports, Urban Mobility, Supply Chain Networks – demonstrated how these technologies can be replicated in more locations.

At the beginning of the project a team of experts together with the project partners proposed a list of 134 Key Performance Indicators (KPIs) across all pilots according to their specific characteristics and requirements. These KPIs were then catalogued under six different assessment categories in line with European Commission best practices: operational efficiency, asset management, environmental quality, energy consumption, safety and economic viability. A four-level assessment was then carried out to classify the impact of the Big Data on the different transport modes from different perspectives and levels (see table):

- **Pilot** level: analysing Big Data impacts for each pilot by aggregating within one of the six assessment categories the improvements related to the specific KPIs

  - **Domain** level: considering only shared or similar KPIs mutually selected by both pilots within the same domain to identify Big Data impacts on a single transport mode by itself

  - **Transversal** level: considering all the KPIs which are generally useful for all the pilots in order to conduct a horizontal analysis across every transport mode for each one of the six assessment categories

  - **Strategic** level: jointly evaluating only the most important KPIs chosen by each pilot to gain overall insights from a more business-related perspective

The formula applied across the evaluation process relies on comparing the KPI value obtained at two clearly separate time slots: “Baseline” stage without leveraging Big Data technologies and “with Big Data Implementation” stage, where new Big Data technologies are leveraged. The percentage of improvement of the KPI is then measured. This is quite a simple formula that helps yield a general overview of the project’s performance when factoring in different transportation modes and variables at the same time.

The best results among the six assessment categories for all pilots are related to operational efficiency and asset management, achieving around 60 % and 50 % improvement respectively.
TransformingTransport data portal

What

The TransformingTransport data portal (https://data.transformingtransport.eu/) is the gateway to the TransformingTransport digital ecosystem. It provides access to all the metadata about the data assets that organisations in TransformingTransport generate and use in the project’s pilots.

Why

When organisations work on data-intensive problems (analytics, machine learning, etc.), they normally use data assets owned and managed by one or several departments, generating new data assets as a result of these processes. However, such data assets are normally not systematically registered or catalogued in a company’s data catalogue. This makes it difficult for any Chief Data Officer or Chief Technology Officer of the related organisations to understand the wealth of data and implications that data management and maintenance decisions may have on the organisation. Furthermore, the rules for data use and sharing are not always clearly specified or understood by people who use these data assets within the organisation.

The challenge is even more relevant in digital ecosystems, i.e. when data from other organisations are also being considered and used in data-intensive tasks. To illustrate, open data from public organisations may need to be included or commercial data may need to be acquired by the organisation under specific licenses and used in combination with open data or the company-owned data. This shows the need for a shared, explicit catalogue of data assets used in data-intensive processes run by an organisation. It calls for better and more systematic management of those data assets and better auditing of all the legal and ethical implications related to using them.

How

The data portal follows the usual practices put in place by public administrations in the release of their open data portals. It uses widely deployed open source technology – i.e. the Comprehensive Knowledge Archive Network (CKAN) – to manage the data asset catalogue, allowing organisations to register and maintain the metadata descriptions of their data assets (whether they’re open or closed).
Ensuring data quality in Big Data

Data quality in Big Data projects is a topic of major importance. Assessing and setting the initial data quality status of a project is fundamental for designating a quality baseline that can be used as a starting point. It is also pivotal to better track the selected data quality indicators and understand variations over time.

The TransformingTransport project has been working since the outset on data quality and has produced recommendations to properly tackle data quality issues in the transport domain. Examples of these TransformingTransport guidelines involve starting from a solid understanding of the operational context and business requirements, as well as focusing on the key data quality indicators that are the most important ones for a specific use case.

In following these guidelines, all the pilots were provided with a common format for describing their data sets (i.e., the Data Asset ID Card template) to enable a data quality assessment process based on two main milestones called ‘metadata sieve’ and ‘data sieve’. This is illustrated in the image below. In order to monitor the overall data quality process, a set of Basic Data Quality KPIs was defined based on the concepts of ‘available’, ‘understood’ and ‘usable’ data assets.

Within this context, the TransformingTransport project collected a set of lessons learned about data quality, representing the bottom-up experience developed so far by the pilots while operating in a challenging environment such as the project’s seven transport sector domains. The key lessons are:

- Data availability is still a challenge but can be tackled by creating digital environments where organisations could publish their data assets with standard descriptions.
- Analysing data quality requires intensive efforts. Being able to leverage both clear understanding of the project use cases and rich data assets descriptions is essential for properly selecting data sets at an early stage and for concentrating efforts related to deep data analysis on a limited number of data assets. This is often crucial for meeting project deadlines.
- Managing data quality is difficult without proper data governance processes, which should be supported by integrating within the data definitions non-technical aspects such as data ownership, licensing schemes, terms and conditions, etc.
- Machine learning can be used to assist data engineers in analysing and improving data accuracy.
- Clustering data according to predefined characteristics that are relevant for the specific domain under investigation helps to gain better understanding about the data and enables the use of sharply defined techniques to conduct data quality analysis.

![Data Quality Assessment Process](image-url)
Legal Notice

The information provided in this publication reflects only the views of the TransformingTransport consortium. The European Commission is not responsible for any use that may be made of it.