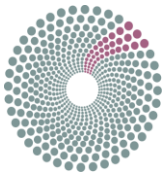

Big Data Europe for

Smart, Green and Integrated Transport

1st Workshop Report

In the framework of the BigDataEurope (BDE) project, ERTICO-ITS Europe organised a workshop on "Big Data for Smart, Green and Integrated Transport" on the 7th October 2015, collocated with the ITS Congress in Bordeaux.

Background

BDE will undertake the foundational work for enabling European companies to build innovative multilingual products and services based on semantically interoperable, large-scale, multi-lingual data assets and knowledge, available under a variety of licenses and business models.

The Transport Societal Challenge set its eyes to contributing to the following objectives: resource-efficient transport that respects the environment; better and more informed door-to-door mobility; less congestion and unforeseen delays; safer and more reliable multi-modal mobility; secured exchange of personal information; global leadership for the European transport industry; and socio-economic and behavioural research and forward looking activities for policy making.

Objectives

Participants to this workshop had the opportunity to influence the design, and ultimate benefit from the Big Data platform that BDE will deliver. This platform aims to facilitate Big Data usage in real world examples, and will consist of an architecture, components, guidelines and best practices to make the best of Big Data in this case in the setting of transport. The workshop focused on the elicitation of requirements for Big Data management within the intelligent transport domain, by encouraging European stakeholders to participate in the definition of Big Data needs and requirements for the intelligent transport domain in Europe.

General information

Over 30 participants attended the workshop, which was more oriented to road transport rather the entire domain. The workshop hosted speakers from various backgrounds be it policy, industry, research institutions or universities; who presented on a wide range of topics ranging from the

role of social media in transport, over open logistics and traffic management to data analyses techniques.

The workshop consisted of three sessions dedicated respectively to data-centric initiatives in transport, big data use-cases in transport, and technologies and tools used and envisaged. Talks in these sessions were followed by three breakout sessions; dedicated to policy, business, and data and technology respectively. The workshop was concluded following a summary of discussions and results in the three parallel sessions .

Agenda

9:00	**Official Welcome**	Maxime Flament, ERTICO – ITS Europe
	Opening Session	
9:15	Big Data for Transport – Why Does it Matter?	Paul Timmers, European Commission
9:30	What can big data do for transport?	Simon Scerri, Fraunhofer IAIS
	Session 1: Data-centric Initiatives in Transport	
9:40	Big Data Sources For/From Intelligent Road Transport: An overview	Evangelos Mitsakis, CERTH
9:50	The use of big data for public transport performance measurement	Roberto Baldessari, NEC
10:00	The role of social media in transport	Dave Marples, Technolution
10:15	***Coffee break***	
	Session 2: Big Data Use-cases in Transport	
10:25	Open Logistics	German Herrero, ATOS
10:35	Real-time Traffic information and traffic management	Nick Cohn, TomTom
10:45	How to deal with all the data that is generated by the vehicle sensors?	Sean Gaines, Vicomtech
	Session 3: Technologies and Tools used and envisaged	
10:55	Existing tools and technologies used across other sectors	Simon Scerri, Fraunhofer IAIS
11:05	Data Analyses in Transport	Helena Gellerman, Chalmers University

	Breakout Sessions
11:20	Leader of the **Business** group: Dave Marples, Technolution Leader of the **Policy** group: Giselle Roesems, DG CONNECT Leader of the **Data & Technology** group: Sean Gaines, Vicomtech
12:30	Breakout session reports
12:55	**Closing note**

Data-centric Initiatives in Transport

The first of three sessions dealt with data-centric initiatives in transport. The use of big and open data in the transport sector is relevant for governments (traffic control, planning and modeling, route planning, congestion management, etc.), for the private sector (travel industry, route planning and logistics, competitive advantages, etc.) and for individuals (route and travel planning).

Big data plays an important role in how smart cities obtain their transportation targets, meaning how smart cities use and deploy ICT to enhance their transportation networks. Big data in transport will lead to improved multi-source traffic and travel data availability and processing, and to tools to enhance multi-source traffic and travel data fusion for, for instance, improved traffic and mobility management. Combining big, open and linked data will leverage innovation and economic benefits.

Big data is important for traffic management: it provides new insights into traffic patterns, real-time traffic data to information service providers. In public transport, bid data helps to understand travelers' journey patterns, informing transport agencies and operators on how different social groups use the public transport system. Big data for private travelers enables personalised information (such as delays). Finally, big data can also be used for asset maintenance offering new opportunities to identify problems more quickly and reduce costs.

Big data is highly valuable to monitor of public transport performance. Currently data analytics use public transport Key Performance Indicators such as Excess Waiting Time (EWT) in order to measure the contractual performance of the public transportation companies. They reveal the worst performing routes, key bottlenecks on the routes, causes for dwell time at bottlenecks and time table improvement margins. Based on this, bus operators and municipalities/authorities can have an EWT profile based on their public transit scores. Simple analytics derive hot spots to attack in order to reduce EWT.

Social media play an important role in transport with continuous, inbound information (stimulus to traffic control) and outbound information (stimulus to road user). Open questions relate to how to decide which social media inflows are reliable and to what degree, what degree does the fact that social media sources are electively published give users open rights to use them, how to

prevent spoof outflows from confusing or malicious users, and how to prevent beneficial services encouraging inappropriate use of social media apps.

Big Data Use-cases in Transport

The first use-case discussed in this session was *Open Logistics*. The use of big data can lead to efficiency gains in the logistics chain. Manufacturers are looking at big data as a catalyst for greater collaboration, enabling more complex supplier networks that focus on knowledge sharing and collaboration as added value. Big data and advanced analytics are being integrated into optimisation tools, demand forecasting, integrated business planning and supplier collaboration & risk analytics at a quickening pace. Big data also contributes to the delivery process.

The second use-case considered was *Real-time Traffic Management*. Big data allows warnings ahead of jams, HOV lane handling, traffic flow predictions, etc. Going from big data to active traffic management requires merging big data with data from fixed sources. The use of archived data allows to improve individual route planning, to measure bottlenecks and delays, to measure system reliability, to determine priorities for infrastructure improvement, and to analyse the impact of the investments made.

The third use-case focused on how to deal with the data generated by vehicle sensors. Data can come from external and internal cameras each with 360° viewing. Data fusion beyond 720° viewing allows Holistic Driving Models, Personalised Driving Models, HMI response, Continuous Monitoring, etc.

Technologies and Tools used and envisaged

In the context of data analysis in transport, the tools and process in the Field Operational Test (FOT) analysis platform were presented. FOTs generate complex data, subjective and contextual data. As a next step, data is processed using HPC and is hosted in databases and files for video. A future big data platform must allow real-time data analysis (instead of research), include visualisation tools that allow data mining and analysis results, automatic video coding data sent over GPRS/Wi-Fi to cloud based storage, efficient data structures to allow efficient data extraction, and will need to support open repositories with high quality context information.

The ultimate goal of the BDE is to design and build a platform that can benefit not only the transport domain but also a variety of other societal challenges. Therefore the presentations of the workshop were wrapped up by Fraunhofer, as project coordinator. Big data involves the analysis of historical data, analysis of actual data with low latency in “real-time”, and interactive analysis by online queries. The question is how this can all be put together in one big data management system. Fraunhofer presented a blueprint of the data aggregator platform that follows the typical Lambda architecture, and utilizes existing solutions such as BigTop components.

Breakout sessions

The three sessions were followed by three parallel breakout sessions focusing on policy, business and data and technology. Each breakout session saw the participation of between 5 and 12 participants.

Policy

The Policy session discussed privacy issues, the role of the public authorities to make data available to the service providers, issues regarding the use of data (how to allow to link and dig into these data), and the public/private role regarding standardisation. If a common standard will be developed, it should be a non-discriminatory standard with open APIs. Policy makers should also provide clarity on the re-use rights of data. The session discussed the issue of contract obligations and whether guidance from policy makers is required to bind a user to one single service. Participants agreed that there should be a "free flow of data initiative" in the EU. The EU should also promote the use of open data. The session discussed the public-private role with regard to data centre regulations: should the market decide or should there be a service level agreement. The policy sessions discussed the tools needed and better guidelines, on how to anonymise and aggregate data; better guidance, understanding and control of big data; principles of privacy design, and a checklist for testing if privacy for design is respected. The EU should also provide guidance on whether the transport sector should have a transport-specific privacy framework or keep the general framework.

Business

The Business breakout session discussed how current businesses are challenged by big data. An issue is to have the right mindset to make data available in the first place. Making data available also involves a risk factor. An education/outreach process to the public at large is required. Through use-cases, people must be made aware that good things can happen. The process will require a political leap of faith and brave policy makers. It is also important to publicise the fact that specific data is available. Business models will arise from those who can turn data into usable information. The market must become better at turning data into information. The quality of the output is important: the higher quality the higher the market value of the data. Current businesses will be challenged by big data. Ultimately the user will choose which services will survive.

Data & Technology

The Technology break-out session discussed the cross-domain nature of transport data use-cases, which involve psychologists, road engineers, vehicle engineers etc. Discussions revealed that the BDE project's role is to quiz the societal challenge domains, including the transport sector, to propose technical solutions without their clear understanding of the underlying problem. The results will be converted to architectural and design requirements. The session discussed the three transport data dimensions and data types including *Infrastructure* (dynamic maps/attributes, maintenance, operational, foresight), *Vehicle* (location, driver monitoring, performance and telemetry) and *Users* (events, state, behaviour and personal activity). Each of these dimensions and data types were evaluated against the three core Big Data V's: *Volume, Velocity* and *Variety*. The results are summarised in the table below. Open issues relate to unforeseen events, diverse data formats, common understanding (lack of), and scalability / volume issues.

		Volume	*Variety*	*Velocity*
Infrastructure	Dynamic Maps/Attributes	+	-	+
	Maintenance	-	-	+
	Operational	+	+	+
	Foresight	?	?	?
Vehicle	Location	+	-	+
	Driver Monitoring	0	0	+
	Performance	+	+	+
	Telemetry	+	+	-
Private Users	Events	?	+	+
	State	-	-	+
	Behaviours	-	+	+
	Personal Activity	-	+	+

Conclusions

This workshop was organised in order to elicit and better understand the requirements that the Transport Societal Challenge poses on a future Big Data infrastructure in Europe. This is not always straightforward since frequently, people involved in the sector do not even have a complete understanding of the big data challenges, and therefore struggle to identify ideal solutions. The outcome of the workshop will therefore feed the requirements elicitation phase of the BDE project, ahead of developing the first prototypes and the selection and initiation of pilots that are to be implemented. The workshop results indicate that there is a clear need for Big Data solutions in Transport, but that there are also very diverse areas for application. Moreover, this first workshop mainly focussed solely on road transportation. At the same time, it was evident that stakeholders are asking some questions which remain unanswered. This confirms that there is a need for more Big Data expertise to contribute to the transport fields with guidance and best

practices. There was also an emphasis on understanding how to better control and measure the quality of the information.

In view of the foreseen Big Data architecture and implementation, it is important to eventually propose generic tools and expertise which will answer a broad scope of transportation topics from logistics, public transport to traffic management. A basis is to convince the transport stakeholders to contribute to the creation of large pools of well-documented and accessible road data, i.e., open and with known velocity, volume and variety. In some sectors within Transport and some regions, this is well understood and we see that opening the data often brings added-value services. By providing expertise, more flexible tools and demonstrations, the BigDataEurope project will help to convince authorities and other players to open their data and focus on building new services for the end users.

One of the major cross-sectoral concerns is the protection of private data, especially any geo-located data. Here, the project could help by outlining best practices of how to deal with privacy protection while preserving the data's value.

Links to other Material

The slides from the various talks are available on slideshare. Photos from the event are available on Flickr. Additional information was reported on the BDE Website and W3C interest group.

Report by Marie de Rycke, Maxime Flament, Andrea Toth (ERTICO-ITS Europe); and Simon Scerri (Fraunhofer IAIS).