LEVERAGING BIG DATA FOR MANAGING TRANSPORT OPERATIONS

Deliverable 4.3 **Roadmap-Research and Policy Recommendations for the Efficient Utilisation of Big Data in the Transport Field (Version 1)**

Jayant Sangwan[1], Christina Nash[1], Tharsis Teoh[2], Ivo Hindriks[2], Minsung Hong[3]

CORTE[1],

Panteia B.V.[2],

Western Norway Research Institute[3]

June 2020

Work Package 4

Project Coordinator

Prof. Dr. Rajendra Akerkar (Western Norway Research Institute)

Horizon 2020 Research and Innovation Programme

MG-8-2-2017 - Big Data in Transport: Research opportunities, challenges and limitations

Distribution level	Public (P)
Due date	30/06/2020
Sent to coordinator	30/06/2020
No. of document	D4.3
Title	*Roadmap-Research and Policy Recommendations for the Efficient Utilisation of Big Data in the Transport Field (Version 1)*
Status & Version	*Final*
Work Package	*4: Horizontal analysis and road-mapping*
Related Deliverables	*D4.1, D4.2 and D4.4*
Leading Partner	*CORTE*
Leading Authors	*Jayant Sangwan, CORTE* *Christina Nash, CORTE* *Tharsis Teoh, Panteia* *Ivo Hindriks, Panteia* *Minsung Hong, WNRI*
Contributors	*Rajendra Akerkar, WNRI* *Isis De Mortel, Bird&Bird* *Cecile Chanel, Bird&Bird*
Reviewer	*Remy Russotto, CORTE*
Keywords	*Big Data, opportunities, case study, technological, policy, legal, ethical, social, economic, political and environmental issues, research and policy roadmap*

Disclaimer

This report is part of the LeMO project which has received funding by the European Union's Horizon 2020 research and innovation programme under grant agreement number 770038.

Executive summary

This deliverable presents a first version of the Research and Policy Roadmap. Building upon the work done under the previous work packages and deliverables 4.1 and 4.2 of the present work package, this deliverable further probes Big Data related challenges and priorities to develop research and policy recommendations.

The research roadmap discusses Big Data research priorities in the transport sector. It does so by looking at the STRIA (Strategic Transport Research and Innovation Agenda) roadmaps, EU data strategy as well as programs/strategies/initiatives of 18 countries both within and outside the EU. It then presents Big Data challenges, based on a review of previous deliverables under LeMO, as well as other relevant Horizon 2020 projects. These challenges have been categorized across 8 standard Big Data dimensions. Lastly, to overcome these challenges and achieve transport related objectives, the roadmap proposes 31 research interventions. A prioritization of these research interventions will be carried out in the next task to complete the road-mapping exercise and present final recommendations.

The policy roadmap uses the 129 policy related challenges & barriers, identified under deliverable 4.1, and bundles them into 22 condensed issues to conduct a survey. The outcomes of the survey are then used to rank these issues and develop a long list of recommendations based on 114 interventions discussed in deliverable 4.2 of LeMO. In the next phase, these recommendations would be prioritized on the basis of a validation workshop with LeMO advisory and reference group experts and a final short list of recommendations will be produced under the next task.

Table of contents

List of Figures

List of Tables

Glossary

Abbreviation	Expression
API	Application Programming Interface
ATM	Air Transport Management
BDA	Big Data Analysis
BDV	Big Data Value
BDVA	Big Data Value Association
CAVs	Connected and Automated Vehicles
EC	European Commission
EEA	European Economic Area
EU	European Union
GDPR	General Data Protection Regulation
GECKO	Governance principles and mEthods enabling deCisions maKers to manage and regulate the changing mObility systems
INEA	Innovation and Networks Executive Agency
IDPs	Industrial Data Platforms
ITS	Intelligent transportation system
LeMO	Leveraging Big Data for Managing Transport Operations
NIS Directive	The Directive on security of network and information systems
NOESIS	Novel Decision Support tool for Evaluating Strategic Big Data investments in Transport and Intelligent Mobility Services
O-D	Origin and Destination
PDP	Personal Data Platforms
REA	Research Executive Agency
SME	Small and Medium Enterprise
STRIA	Strategic Transport Research and Innovation Agenda
TRIMIS	Transport Research and Innovation Monitoring and Information System

1 Introduction

1.1 Abstract

In this document we look at the research and policy roadmap to foster the use of Big Data in the transport sector and promote sustainable transportation. Building upon the work done in the past work packages, both research and policy sections of the roadmap have discussed various challenges and issues in Big Data and Transport and have provided a list of interventions and recommendations.

However, this is a preliminary version of the roadmap and is only a step towards the development of final prioritized list of interventions and recommendations. To this end, a validation workshop with a group of experts is foreseen as a next task.

1.2 Scope, methodology and structure

The scope of the roadmap is to identify a range of interventions and recommendations that can be used by researchers and policymakers to address specific challenges, support the use of Big Data technology & foster the growth of data economy in the transport sector.

The roadmap is the result of work done in the previous phases of the LeMO project. Under work packages 1 to 3, literature review and case studies were conducted, which were used by the current work package (i.e. work package 4) to identify a set of key issues and challenges (see deliverable 4.1), as well as various opportunities and interventions (see deliverable 4.2). Under the present deliverable, this past work has been consolidated and organized as a step towards prioritizing the challenges/issues and developing final recommendations based on such priorities. Figure 1 provides an easy understanding of the steps to develop the roadmap.

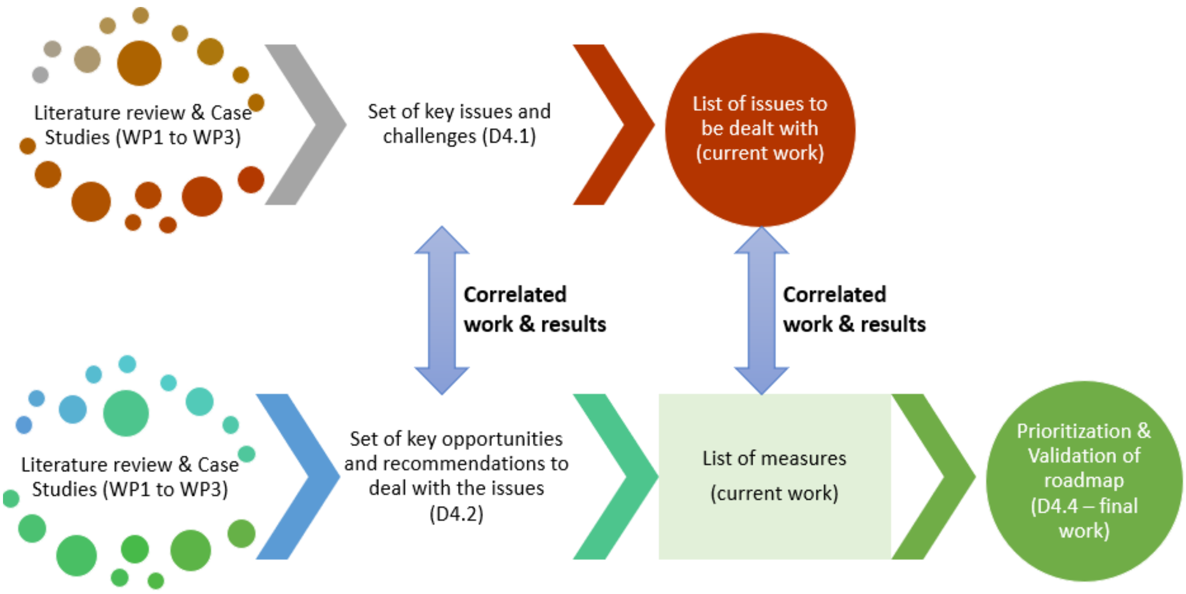


Figure 1 Methodology for developing roadmaps

For the ease of appreciation by different audience groups, the deliverable has been divided into two separate sections, one focusing on research and the other on policy. Accordingly, the challenges and interventions discussed in each section should be construed by the reader in the context of the section within which they appear. Nonetheless, there are methodological overlaps between the two sections, which connect their outcomes. This will be especially visible in a final version of the document, where the prioritization of interventions/recommendations would be conducted by relying on a shared ranking of challenges.

1.3 Target audience

The target audience for this deliverable is:

- Partners and Advisory & Reference Group in the LeMO project;
- European Commission;
- EU Parliament;
- Horizon 2020 projects and related transport projects (cfr. clustering activities);
- Organisations and experts involved in the LeMO case studies;
- Public and private transport organisations;
- Authorities (regional and national level) that develop and enforce policies and legislations;
- EU research agencies, INEA and REA, as well as national research agencies and associations.

2 Research Roadmap

Building upon the work done in the previous work packages, sections 2.1 to 2.5 below constitute the first version of the Research Roadmap in the LeMO project. They include:

1. An overview of the Big Data research priorities in STRIA roadmaps, along with an analysis of the state of play in Big Data & Transport related research;
2. A comparative analysis of programs/strategies/initiatives across European Union and eighteen (18) countries ("countries") from within and outside the EU;
3. A discussion on the challenges in Big Data domain that could be addressed through Big Data research, categorized across eight (8) dimensions;
4. 31 possible research interventions that could help achieve different transport objectives and overcome Big Data related challenges.

Lastly, Section 2.5 includes tables that describe linkages between the *priorities* (identified through STRIA and across different countries) and the *Big Data challenges* (across different Dimensions). This analysis will be further refined in the next phase where results from the survey on the challenges as well as the outcomes of the validation workshop would be used to enrich the results.

2.1 *Transport objectives that benefit from Big Data research Interventions*

This section delves into the EU transport objectives that can benefit from Big Data Research. Since digitalization is more strongly associated with the use of Big Data in transport, the first subsection provides a magnified view on what digitalization entails for transport in Europe and considers what type of research priorities can be identified in this context, including through the use of digital data. It primarily focuses on linkages between Big Data research and STRIA roadmaps.

The second subsection maps the 'transport challenges' that researchers are trying to solve through Big Data research and considers their perceptions on the future directions this research will take.

2.1.1 Broad EU transport objectives, digital transformation, and role of Big Data research in STRIA

Broadly, the European goals of a competitive and sustainable transport system, as set out in the 2011 White Paper[1] are intended to be achieved through the establishment of a single European market; fostering innovation for new technologies and promoting sustainable behaviour; as well as by focusing on modern infrastructure, smart pricing and funding.

However, to account for changes in the recent years, especially technological advances, and the adoption of the EU Green Deal,[2] the development of a comprehensive strategy on sustainable and smart mobility has been announced.

[1] https://ec.europa.eu/transport/themes/strategies/2011_white_paper_en
[2] https://ec.europa.eu/commission/presscorner/detail/e%20n/ip_19_6691

Though the new strategy is expected to come out in the 4th Quarter of 2020, some insights can already be found on its major themes.

In the Graz declaration[3], which was a joint statement made by EU transport and environment ministries; the following core objectives for the transport sector are identified:

1. Clean vehicles: rapid introduction of zero-emission vehicles and decarbonised fuel options,
2. Strategy for sustainable mobility management and planning,
3. Active mobility to promote health and sustainability,
4. Safe and inclusive mobility, and
5. Multimodality and infrastructure.

In her speech[4] to the European Parliament the Commissioner for Transport has further indicated four areas of actions in the upcoming strategy:

- Boosting the uptake of clean vehicles and alternative fuels for road, maritime and aviation through specific initiatives to ensure the availability of marine alternative fuels and sustainable aviation fuels.
- Increasing the share of more sustainable transport modes such as rail and inland waterways and improving efficiency across the whole transport system.
- Incentivising the right consumer choices and low-emission practices.
- Investing in low- and zero-emissions solutions, including infrastructure.

Digitalization also remains to be a major topic for this strategy, and it is intended to be used for – improving traffic efficiency, reducing traffic hazards, combining transport modes and decreasing reliance on private vehicles among other things.[5]

Since digitalization is more strongly associated with the use of Big Data in Transport, it will be important to have a magnified view on what digitalization entails for Transport and to consider what type of research priorities can be identified in this context, including through the use of digital data.

The impact of digital transformation[6] on transport has been studied in a recently published report by the Joint Research Centre of the European Commission[7]. The report notes that

[3] https://www.eu2018.at/latest-news/news/10-30-Graz-Declaration.html
[4]https://ec.europa.eu/transport/themes/strategies/news/2020-02-03-commissioner-valeans-speech-eu-strategy-mobility-and-transport_en
[5] *Ibid*
[6] Digital Transformation refers to refers to the profound changes that are taking place in the economy and society as a result of the uptake and integration of digital technologies in every aspect of human life.
[7] Digital Transformation in Transport, Construction, Energy, Government and Public Administration available at https://publications.jrc.ec.europa.eu/repository/handle/JRC116179

advances in digital technologies and greater connectivity are causing a paradigm shift in the transport sector. It states that digital transformation is:

- facilitating the emergence of new business models,
- making vehicles smarter,
- changing societal culture from private vehicle ownership to shared and on-demand mobility services, and
- enabling reinvention of old production processes and service delivery methods.

To take appropriate advantage and make better use of digital transformation, innovation, and technologies in the transport domain, in 2017 the European Commission adopted the STRIA[8] as part of the 'Europe on the move' package. STRIA not only highlights the transport research and innovation areas, but also lists the priorities for clean, connected, and competitive mobility.

Under STRIA, seven (7) areas reflecting the impact of digitalization on transport have been demarcated. Roadmaps[9] for each area have also been developed, identifying the key research innovation actions and challenges. These areas include:

1. Cooperative, connected and automated transport
2. Transport electrification
3. Vehicle design and manufacturing
4. Low-emission alternative energy for transport
5. Network and traffic management systems
6. Smart mobility and services
7. Infrastructure

Big data technologies and research can play a substantial role as "enablers" for achieving the above described transport objectives as well as for overcoming transport challenges set out in different STRIA roadmaps. This is especially true for *Cooperative, Connected and Automated Transport*; *Vehicle Design and Manufacturing*; *Network and Traffic Management Systems*; *Smart Mobility and Services*; and *Transport Infrastructure areas*, where Big Data technologies and research can find a more direct application. The table below[10] provides an overview of the different Big Data related research priorities that emerged through a scan of different STRIA roadmaps.

[8] https://ec.europa.eu/info/research-and-innovation/research-area/transport/stria_en
[9] https://trimis.ec.europa.eu/roadmaps

[10] The priorities listed in the table, are not an exhaustive stipulation and are only indicative of some of the key themes. For a more comprehensive overview, domain experts are advised to go through individual roadmaps for each strategic area.

Table 1 Research priorities on Big Data in STRIA roadmaps

STRIA areas	Big Data research focus areas and priorities
Cooperative, connected & automated transport	**Data storage**; **data sharing**; Developing standard models on data value chain; **Data mining**; Data analytics; Use of artificial intelligence and Big Data to improve situational awareness.
Transport electrification	N/A
Vehicle design and manufacturing	Focus on **use of Big Data for** automation in design and **manufacturing process** where complexity, process variability, and capacity restraints are present; Use of Big Data is also considered vital for **industry 4.0** in order **to achieve high quality products quicker** as well as in a more resource efficient and flexible way; Smooth, quick and safe **exchange of data** between stakeholders to eliminate inefficiencies in materials supplies and changing the vehicles' architecture; and Research with regards to **data compatibility**, **interchangeability**, **security** and **data standardization**.
Low-emission alternative energy for transport	N/A
Network and traffic management systems	**Use of Big Data for better optimising traffic flows** is considered a priority. However, one of the key issues identified in the roadmap is to define and implement harmonised and standardised concepts for (cross-border) **exchange of data** between authorities, operators and/or users; to enable an easy and reliable **data synchronisation**; and to interconnect the national systems with (Pan-) European systems. It is considered that **Big Data can be useful** to Network and Traffic Management systems **if the generated data across modes can be fused and managed in an efficient manner** while protecting the privacy and security of goods.
Smart mobility and services	Among other things it is deemed important – to **collect and collate systemic and dynamic mobility data** to contribute to effective policy-making and implementation, identify mechanisms to **ensure fair access to digital infrastructures**, develop **mobility data management solutions**, and support the **development of technical standards for communication and interoperability** of user devices, vehicles, critical infrastructures, energy systems and mobility data.
Infrastructure	Following priorities are seen to emerge in the roadmap: (a) development of smart infrastructure by adapting infrastructure to support development of new technology and by providing full **interoperability** in modal chain to allow seamless switching between modes and across borders; (b) development of connected infrastructure by setting up infrastructure for **collection and management of data**, by creating added value from data collected from physical and digital infrastructure, by **creating connected environments such as logistics hubs and parking lots**, and by identifying infrastructure and equipment required for deploying technologies facilitating autonomous and multimodal transport; (c) **leveraging** digitalization, artificial intelligence and **Big Data for automated construction and maintenance** as well as to manage assets, plan and deliver infrastructure; (d) leveraging automation, robotization and remote solutions for **developing smart, self-monitoring and self-healing infrastructures** as well as to conduct routine periodic maintenance, inspections, emergency works, thereby reducing risks and costs.

2.1.2 Transport & mobility topics in Big Data research and future directions

The application of Big Data in the transport domain is being continuously explored by researchers. The present subsection provides an overview of this work by discussing the outcomes of several papers and projects that have studied the current state of Big Data and Transport related challenges, research and industry applications.

2.1.2.1 Current State

C. Katrakazas et al (Katrakazas, et al. 2019), identified major transportation areas and sub-problems that could benefit from Big Data following research conducted within the framework of the NOESIS project[11], some of which include:

- ***Environment***: where it is suggested that Big Data could help in addressing environmental issues related to transport in areas such as – Decarbonization; Transport electrification; Noise and vibration reduction; Resource Consumption; and Environmental Monitoring.
- ***Connected and Automated Vehicles (CAVs)***: where it is believed that Big Data can support in handling mixed traffic scenarios, development of resilient systems for navigation and localisation, use of mix communications and cooperative levels in ITS technologies, deployment of connected and multimodal transport solutions, development of a framework for human-machine interactions, human- like driving/navigation behaviour, human acceptance levels for automated transport, as well as Implementation or testing protocols.
- ***Safety and security***: where Big Data is considered useful for Human events management, Human factors and human behaviour modelling, safe mobility of challenged passengers, preventing transportation of illicit goods, enforcement and protection against theft or other crimes, surveillance, vehicle safety design, and contingency/recovery planning.
- ***Traffic management and operations***: where it is suggested Big Data can help areas such as traveller information systems, resources optimization and cost management, demand and delays management, parking demand management and modelling, traffic management in emergency or extreme events, management of traffic environment changes e.g. MaaS, automation, slot allocation; congestion pricing, short-time forecasting.
- ***Transport planning and policy***: where it is found that Big Data can play a role in optimization of the regulation/privatization levels, transport sustainability, governance, social justice including intergenerational equity, impact assessment and ex-post evaluation of measures/treatments and integration of land use planning and environmental concerns.
- ***Freight and logistics***: where Big Data can help in aspects such as supply chain management, vehicle capacity optimization, transparency and traceability of goods,

11 https://cordis.europa.eu/project/id/769980

inter-modality/ synchro-modality, driver shortage and retention, speeding up delivery times; pro-active operational prediction and route programming.

- ***Railways**:* In the railway sector, it is considered that Big Data can help in information management (e.g. passenger information, ticketing, operation management and tracking), train control (e.g. balise data, communication systems, automation, localisation), energy (smart metering and intelligent power supply), infrastructure (monitoring, surveillance analytics, track condition, signalling systems, emergency communications) and predictive maintenance (real-time re-scheduling, rail decision support systems, safety).
- ***Aviation**:* In the aviation sector it is suggested that Big Data can help in Operations and Air Transport Management (ATM) (mitigation of delays, weather resistance, optimization of flight plans), security and safety (e.g. on-board monitoring, all weather same airspace operations, privacy, resilience against cyber-attacks), integration of new technologies (e.g. drones in air logistics), passenger centric ATM and security solutions, as well as aircraft trajectory prediction, air traffic scheduling and routing.

In addition to the above, the paper also suggests that Big Data can provide several benefits for *maintenance* and *increasing cost efficiency* in the transport sector.

T.F. Welch and A. Widitia have reviewed sources and methods of Big Data in Public Transport (Welch and Widita 2019). While analysing eighty-one (81) studies on Big Data and transport they have grouped the topics considered in these studies under the following categories:

- ***Service and performance***: covering uses of Big Data that can help agencies evaluate their services and identify potential improvements e.g. improvement of traffic efficiency, traffic information service, transport structure, energy and pollution reduction and social inclusion.
- ***Transit users' behaviour***: related to better understanding (individual and aggregated) user behaviour over space and time as well as due to influencing factors such as weather and disasters, for improved decision making.
- ***Travel demand***: related to calculation of travel demand especially focusing on Origin and Destination (O-D) matrices and modelling transportation mode split for better public transport planning.
- ***Management***: focusing on methods allowing different agencies to maintain infrastructure, manage and use large amounts of data, from diverse sources.
- ***Resilience and health/safety***: covering resilience of transport infrastructure and impact of transport systems on public health and safety e.g. recovery of transport system after natural disasters and identifying unsafe transit points.

Another interesting finding from the paper suggests that the most common topics studied by researchers relate 'Travel Behaviour', 'Travel Demand' and 'Service' (approx. 75% of the papers address these topics). Also, the most common methodologies employed by researchers include 'descriptive – data visualization', 'estimation' and 'regression' (Welch and Widita 2019).

Similarly, K.E Zannat and C.F. Choudhury have also reviewed the current state of art of Big Data in public transport planning (Zannat and Choudhury 2019) where they consider more than forty-seven (47) relevant papers. The topics considered in these research papers have been organized in the following categories:

- ***Theme 1*** - *Use of Big Data in travel pattern analysis:* covering use of Big Data to support urban planning by understanding travel behaviour. Here identifying aggregated and individual travel behaviours along with integration of socio-demographic attributes were identified as key topics. 36% of the papers addressed this theme.
- ***Theme 2*** - *Use of Big Data in public transport modelling:* focusing on use of Big Data in trip generation[12], trip distribution[13] and trip assignment through the estimation of Origin and Destination (O-D) and route-choice modelling. 30% of the papers addressed this theme.
- ***Theme 3*** - *Use of Big Data in public transport performance assessment:* related to measuring aspects such as waiting times, arrival times, impact of fare policy etc. using Big Data, with the aim of improving service quality.

In addition to the above papers, research conducted within the framework of LeMO project as well as the results of a survey conducted within the framework of GECKO[14] project are worth mentioning.

In deliverable 2.1. of the LeMO project[15], the following have been identified as Transportation areas that can directly Benefit from Big Data research and technology:

Table 2 Direct improvements to the transport system through the use of Big Data

1.	**Shared situational awareness**
2.	**Improving transport network capacity** • Long term planning of the transport network • Realtime traffic management • Mitigating risks that degrade network capacity
3.	**Improvement of transport services** • Marketing activities in the transport sector • Vehicle routing • Service provision and schedules
4.	**Shift towards sustainable mobility and transport**

[12] Trip generation is the process of estimating the number of trips that will begin or end in each zone within a study area.

[13] Trip distribution is the process by which all the trips generated in a study area are allocated among the zones.

[14] GECKO refers to Governance principles and mEthods enabling deCisions maKers to manage and regulate the changing mObility systems. It has received funding from EU's Horizon 2020 research and innovation program. For more information see http://h2020-gecko.eu/

[15] D.2.1 of LeMO project (https://lemo-h2020.eu/newsroom/2018/8/31/deliverable-21-report-on-economic-and-political-issues)

In the GECKO project, the results of a recently conducted survey with industry stakeholders reveal areas and purposes for which 'new mobility service providers' are applying Big Data:

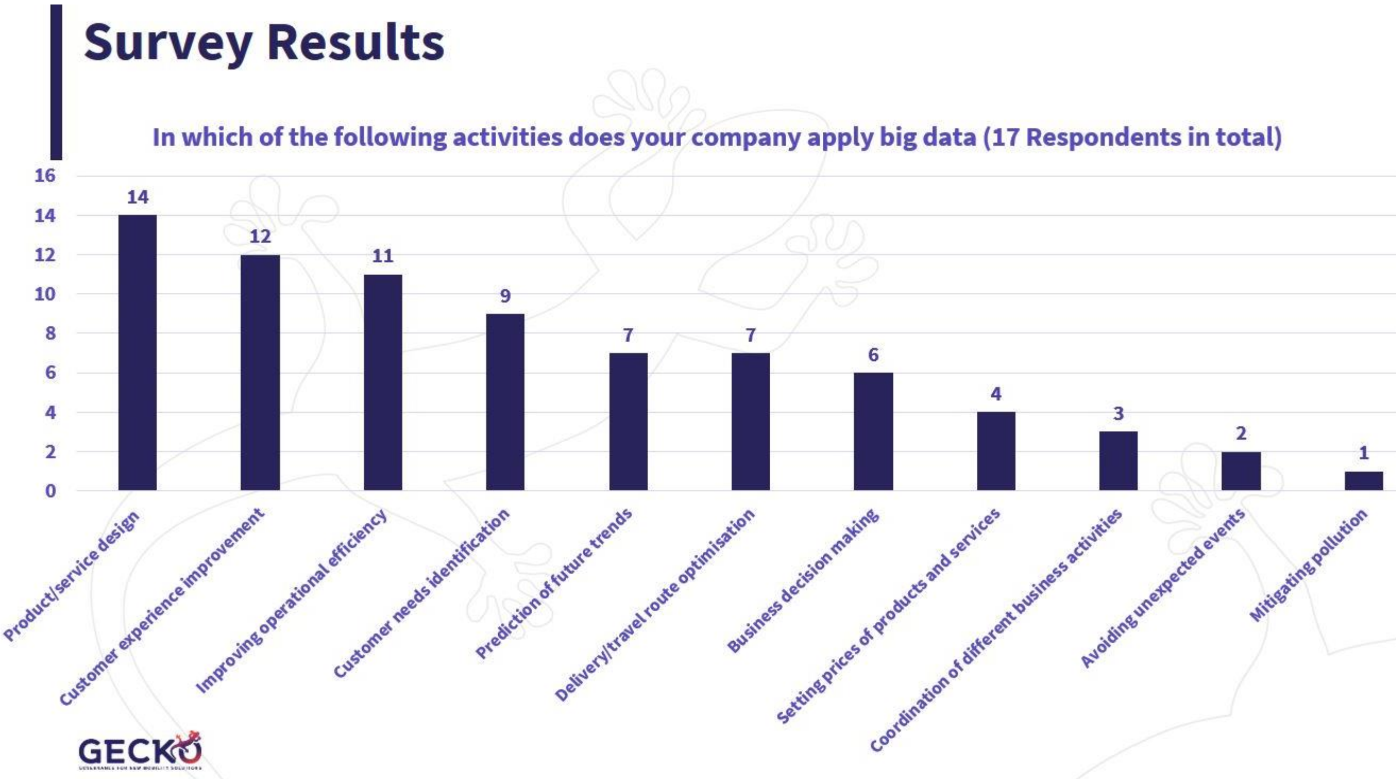


Figure 2 Survey results from GECKO project

2.1.2.2 Future Direction

A.Neilson et al, while discussing the state of research on Big Data in Transport (Neilson, et al. 2019), have indicated that the state of Big Data in Transport is still not mature, as many research projects are limited in their nature, conducted at university level and are of a limited scale. Given the complexity of transportation systems, they suggest deployment of Big Data systems through large (citywide) and multi-year (five years or longer) projects in coordination with a range of stakeholders and departments in a city.

According to (Welch and Widita 2019), studies on: effectively managing and utilising the large amounts of data available to transit agencies and decision-makers, the relationship between bike shares and rail and bus transit by using Big Data from rail and buses combined with bike share data, combining Big Data from public transit sources with data related to automobiles; resilience and health/safety in the public transit realm and on methods to easily process data from multiple sources and provide user-friendly output– could prove useful in the future owing to a current lack of such studies and due the impact these studies can have on public transit systems as well as on society.

In (Zannat and Choudhury 2019), K.E Zannat and C.F. Choudhury point out that there is a research gap in applying Big Data techniques to new areas in transport domain, accordingly they suggest development of more dynamic planning models as well as the development of new techniques to validate the underlying assumptions in Public Transport modelling and to provide insights about their forecasting performance. They suggest that there are possibilities of combining multiple Big Data sources or big and small data for improving Public Transport planning. While there are various sources of Big Data, 'methods' to integrate these different types of data have not been fully explored. Doing so could help solve problems such as congestion and service deficiency as well as improve accuracy in individual travel behaviour. In addition to providing a better understanding on travel behaviour, improved data quality and other data related aspects, could foster a better understanding of public transport planning by considering multimodal data.

On a higher level, the direction that future research would take is closely related to the larger policy objectives of the EU and its Member States. Application of good governance through appropriate regulatory frameworks on use and access to data in a secure manner, while ensuring privacy, will also be crucial for achieving these objectives and determining how the future of using Big Data in transport unfolds.

The next section looks at how Big Data research features in the programs/strategies/initiatives of different countries.

2.2 Comparison of research direction in different countries and the EU

This section considers programs/strategies/initiatives of the EU and several different countries from within and outside the EU. It provides a high-level overview of different countries' perception of the role of 'Big Data research' in the Transport sector.

Later, in Section 2.5, the programs/strategies/initiatives identified below are linked to the various Big Data challenge Dimensions discussed in Section 2.3, to gain an insight on which Big Data challenges are more prevalent in different countries. Though this is not an exhaustive analysis, nevertheless its relevance cannot be understated, especially in prioritizing those Big Data Dimensions which are commonplace in many countries.

2.2.1 European strategy for data[16]

The European Commission (EC) is a lead role model for a society empowered by data to make better decisions and this vision stems from European values and fundamental rights. With the strength of the Single Market's regulatory environment, the EC is strongly interested in leading and supporting international cooperation with regard to data (such as China and the US), shaping global standards and creating an environment in which economic and technological development can thrive, albeit in full compliance with EU law.

The EC sees common European rules and enforcement mechanisms which allow data to flow, while respecting personal data protection and consumer protection legislation as a key step to achieving this vision.

[16] https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

The EC has identified the following challenges, some of which overlap with the challenges identified under LeMO and which will be discussed in detail in Section 2.3:

- Availability of data
- Imbalances in market power
- Data interoperability and quality
- Data governance
- Data infrastructures and technologies
- Empowering individuals to exercise their rights
- Skills and data literacy
- Cybersecurity

The EC's strategy is to propose policy measures is based on four pillars:

Table 2 Four pillars in the European Commission's strategy

Pillar 1. A cross-sectoral governance framework for data access and use:	This entails the greatest degree of direct regulatory intervention such as a framework for the governance of common European data spaces (Q4 2020), an implementing act on high-value data sets (Q1 2021), explore the need for legislative action on issues that affect relations between actors in the data-agile economy (*possible new Data Act 2021).*
Pillar 2. Enablers: Investments in data and strengthening Europe's capabilities and infrastructures for hosting, processing and using data, interoperability:	High impact projects on European data spaces, sign MoU with MS on cloud federation Q3 2020.
Pillar 3. Competencies: Empowering individuals, investing in skills and in SMEs (explore enhancing the portability right for individuals under Article 20 of GDPR)	Giving individuals more control over their data through tools and means, such as: - enhancing GDPR - Digital Europe programme. By 2025, EU should half the current gap of 1m digital specialists. Dedicated capacity building for SME's.
Pillar 4. Common European data spaces in strategic sectors and domains of public interest	To complement the horizontal framework (et al), EC promotes the development of common European data spaces in strategic economic sectors and domains of public interest. EC will build on work with research community with the European Open Science Cloud and will support the establishment of 9 common European data spaces.

In addition to a general framework for a cross-sector data space, the EC also intends to promote the development of European data spaces in **strategic economic sectors** and domains of public interest. The intention is to make available large pools of data in these sectors and domains along with the technical tools and infrastructures necessary to use and exchange data and appropriate governance mechanisms. There are seven sector-based data spaces identified including a **common mobility data space** - to facilitate access, pooling and sharing of data from existing and future transport and mobility databases.

2.2.2 China

In 2006, China issued '*The 13th five-year plan for economic and social development of the People's Republic of China (2016-2020)*'[17] (five-year plan), which sets forth China's strategic intentions and defines its major objectives, tasks, and measures for economic and social development. Whilst every conceivable sector of society [18] has been earmarked for development in the far-reaching five-year plan; transport[19] has also been identified in the five-year plan in which **Big Data's pervasiveness features prominently** and is referred to in at least nine chapters of the five-year plan. **Transport** is mentioned in the context of emerging industries, transportation cities, harmonious and pleasant cities, regional development, special regions etc.

Separately, an academic report entitled '*Big Data analytics for transportation: problems and prospects for its application in China*' (Robert, Weng and Simon 2016), sheds light on some challenges and opportunities associated with intelligent transportation in China. In terms of challenges, these include: China's large transportation systems and saturated usage, city structure and law enforcement and sensitivity to special events. The areas of opportunity identified in the report are interesting and are inter-connected: China's ability to develop standards for its own Big Data systems, for e.g. by a national standards body, the latter opens the door to vendors of Big Data software systems components, which in turn will feed into building a national transportation Big Data analytics infrastructure.

2.2.3 United States of America

In 2016, the U.S. Department of Transportation (DOT) published a report '*Research, Development and Technology (RD&T) Strategic Plan 2017-2021 ('Strategic Plan'*)[20]. Big Data is identified as one overarching research theme that intersects with critical transportation topic

[17] https://en.ndrc.gov.cn/policyrelease_8233/201612/P020191101482242850325.pdf

[18] Chp 6 Science & technology, Chp 16 Financial sector, Chp 20 Agriculture, Chp 25 Information networks, Chp 36 Urban & rural development

[19] Chp 2 Emerging Industries, Chp 29 Transportation cities, Chp 30 Energy, Chp 34 Develop harmonious & pleasant cities, Chp 37 Regional development, Chp 39 Yangtze, Chp 40 Special regions, Chp 44 Environmental governance, Chp 49 Strategy for opening up

[20] https://www.transportation.gov/sites/dot.gov/files/docs/USDOT-RD%26T-Strategic-Plan-Final-011117.pdf

areas. Altogether these themes include policy research, emerging technology, strengthening research coordination, and Big Data.

U.S. DOT firmly acknowledges that research plays a crucial role in enhancing the techniques by which digital technologies can render data accessible by decision makers in real-time to help make accurate, informed decisions. To this end, it sets out the following objectives:

- Bolster research into Big Data architectures and management techniques
- Discover new analytical applications for Big Data
- Invest strategically in data preservation and use

One focus of these objectives is high-level data analyses of U.S. transit system safety to identify critical issues and emerging issues and trends in different modes of public transportation. Projects will include enhanced safety data collection and potential applications of Big Data to identify safety issues and enhance safety research. Another report, '*Beyond Traffic: Trends and Choices 2045*'[21] projects a 30-year outlook on the future of the transportation system in the USA, and fully recognises the era of Big Data.

Maine Department of Transportation is currently undertaking a project; '*Current Status of Transportation Data Analytics and A Pilot Case Study Using Artificial Intelligence (AI)*'[22]. It involves a comprehensive review of all the data and data needs that state DOTs possess, relating to traffic operations. In this way, it directly addresses data problems such as different divisions of DOTs collecting data, which may lead to duplication of efforts and under-utilization of data. This has revealed that sometimes one DOT division is not aware of the existence of a data set maintained by another division. The power of the data is not fully understood and exploited as it should be. This is why it is important to conduct this review of data, data needs, data analysis and decision-making practices, with a view to co-ordinating better data collection, reduction, storage, and analysis efforts.

2.2.4 United Kingdom

The UK is very active in maintaining its vision as a world leader in Big Data, as cited in '*Seizing the data opportunity – A strategy for UK data capability (October 2013)*'[23]. There is an abundance of publications regarding Big Data and R&I in the UK. The UK's long-term strategy as regards R&D strategy, included in which is transport, is documented in '*Industrial Strategy – Building a Britain fit for the future (Industrial Strategy)*[24]. The Industrial Strategy is supported by the International Research and Innovation Strategy[25], which sets out how the UK will develop

[21] https://www.transportation.gov/policy-initiatives/beyond-traffic-2045-final-report
[22] https://rip.trb.org/Results?txtKeywords=big+data#/View/1704050
[23] https://www.gov.uk/government/publications/uk-data-capability-strategy
[24] https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future
[25] https://www.gov.uk/government/publications/uk-international-research-and-innovation-strategy

its international research and innovation partnerships to help achieve the targets as set out in the Industrial Strategy.

A policy paper entitled '*The Grand Challenge*'[26] updated on 13 September 2019, identifies the following challenges to transform the future: artificial intelligence and data economy, clean growth, future of mobility and ageing society.

The project '*Decentralised real-time electric vehicle charging: optimality, fairness and resilience (01/17-12/18)*[27]', involves real-world data describing mobility and charging behaviour, which is considered pivotal in understanding the demands on the networks and used in simulation models.

2.2.5 Canada

Canada's '*A National Data Strategy for Canada – Key Elements and Policy Considerations*'[28] identifies the sectors in Canada where a national data strategy is most needed and examines the domestic and international policy considerations. Cities are identified as an area where Canada needs to think beyond open data initiatives when planning their technology strategy. Unless they commit to building their infrastructure with open technology, cities will struggle to fulfil their vision for a smart urban environment.

The transport research policy for Canada is set out in the 2017-2020 '*Federal Sustainable Developmental Strategy*'[29] (Strategy). There are five key theme areas in the Strategy, one of which is to invest in a greener, more innovative transportations sector that embraces new technologies to improve Canadians' lives.

'*Network and asset management: benefits of real-time data (2018)*'[30] is a research programme, which explores the opportunities and challenges of real-time technology and data in supporting network and asset management activities. Its activities included identifying a range of opportunities for expanding the use of real-time technology, for example by making better use of existing datasets and improving the detection of incidents and defects. A range of challenges were also identified, including the difficulties of working with 'big' real-time datasets, the risk of reliance on technology and the need for specialist expertise to develop real-time applications.

---

[26] https://www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/industrial-strategy-the-grand-challenges

[27] http://www.erachair.uniza.sk/project-on-decentralised-real-time-electric-vehicle-charging/

[28] https://www.cigionline.org/publications/national-data-strategy-canada-key-elements-and-policy-considerations

[29] https://www.tc.gc.ca/eng/2017-2020-departmental-sustainable-development-strategy.html

[30] https://www.tac-atc.ca/en/bookstore-and-resources/tac-news/information-bulletin/network-and-asset-management-benefits-real-time-data

2.2.6 India

India's current R&D strategy is set out in the '*Science, Technology and Innovation Policy 2013 (STI)*'[31]. This sets out the vision of STI in India with aspirational and broad-ranging goals. STI is recognises as major drivers of national development globally. It seeks to accelerate the pace of discovery and delivery of science-led solutions for faster, sustainable and inclusive growth.

The Department of Science & Technology's *Annual Report 2018-2019*[32]*,* evaluates how the objectives are being achieved for that year and the measure that must be taken to improve. There are a number of outputs mentioned, such as the initiation of new programme of cooperation with France on smart mobility & transportation and clean transport with Sweden. There are new national knowledge network programmes in place to tackle many challenges attributed to the projected population growth of about 50% of its population, such as urban transportation.

The National Innovation Council (NIC)[33] was established by the Indian Government in 2010 to develop a vision and a road map for promoting innovation in industry, society and the economy. Through the NIC, there are a range of national R&D programmes.

2.2.7 Iceland

The Icelandic Road and Coastal Administration (IRCA) funds the '*Road State Monitoring System (ROSTMOS)'* project, as part of its role to oversee Iceland's transport system on land and sea. ROSTMOS aims to achieve the better use of information technology for the assessment of roads and uses real-time data to assess the condition and winter maintenance action. The Icelandic Traffic Research Board (RANNUM)[34] produces approximately 15-20 different projects each year as part of its purpose to promote research that can be applied to prevent road accidents and to gather new national and international knowledge in the road safety sector.

'*Readiness assessment for autonomous vehicles in Iceland*' (01/15-05/16)[35], is a project that examined the aspects of the environment and other infrastructure on which self-propelled cars rely on ploughing and steering, with a view to assessing whether Iceland can adapt to driverless cars. It reviewed technology from companies in this field such as Google, Audi, Mercedes Benz, GM, Volvo etc.

2.2.8 Switzerland

Swiss Association of Road and Transportation Experts (VSS)[36], is a union of transport sector entities that conduct research for the development and extension of application-oriented

---

[31] https://dst.gov.in/st-system-india/science-and-technology-policy-2013
[32] https://dst.gov.in/about-us/annual-reports
[33] http://innovationcouncilarchive.nic.in/
[34] http://www.vegagerdin.is/upplysingar-og-utgafa/rannsoknarskyrslur/umferd/rannum/
[35] http://www.vegagerdin.is/um-vegagerdina/rannsoknir-og-throun/ransoknaverkefni/alm2015//58c351e63f2888ed00257e0700395fee?
[36] http://www.vss.ch/

expertise in road and transportation. '*Multifunctional small and medium freight facilities (01/15-07/16)*'[37], was a project that analysed the state of the art and best practices regarding multifunctional small and medium transhipment facilities. Loading and unloading technologies formed an important input. The Swiss competence Centre for Energy Research in Efficient Technologies and Systems for Mobility (2014-2016)[38], includes five (5) capacity areas covering two innovation fields for the development of knowledge and technologies for the transition from the current fossil fuel-based transportation system to a more sustainable one. There are innovation fields which deal with components and devices for e-mobility and mobility systems and mobility systems with special focus on infrastructure integration, improved train operation, urban planning, monitoring and user communication.

2.2.9 Belgium

In Belgium, there are several active R&I projects being undertaken in the regions and communities of Belgium involving transport. For example, in Flanders, '*Mobilidata*'[39] is a five-year project from 2019-2023 with an investment of EUR29m, which is part of the Flemish government's vision for a 'smooth and smart mobility system'. It is also a constituent part of one of the Flemish government's priority to provide free-flowing, safe ways of getting around. Big Data plays a prominent role in this project as it attempts to co-create a whole raft of innovative mobility solutions with private partners, based on a digital infrastructure, smart traffic lights and high-quality, sustainable data sources. The project aims to invest in intelligent traffic lights cooperating in the Cloud, develop an open data system enabling road users to access and deliver real time information, develop R&D for the 'next generation' mobility use cases with a focus on vulnerable road users and screening of market readiness of technology.

2.2.10 Austria

'*Take Off 2015-2020'* is a funding program which came into formation after a stakeholder process that was spearheaded by the Austrian Federal Ministry of Transport, Innovation and Technology in 2014, involving participants from the Austrian R&I sector, international experts and representatives of the European Commission. Whilst the broad aims of the program include the building of strategic partnerships, strengthening Austria's competencies and skills in R&I etc; the following has been specifically identified as regards transport: transport electrification, network and traffic management systems, infrastructure etc.

2.2.11 Croatia

Croatia's '*Strategy for Transport Development 2014-2030'* is part of a Partnership Agreement between Croatia and the European Commission, which sets down the strategy for the optimal use of European Structural and Investment Funds. In particular, the '*Operational Programme*

[37] https://www.aramis.admin.ch/Grunddaten/?ProjectID=36189
[38] https://www.sccer-mobility.ch/
[39] https://mobilidata.be/en

Competitiveness and Cohesion 2014-2020[40]', seeks to develop many sectors including that of transport. In terms of Big Data and transport, there is a procurement project for an IT solution for analytics of large data sets of e-mobility.

2.2.12 France

Through France's '*Sustainable Transport and Mobility Programme*', 17 projects have been undertaken encapsulating all land transports, airfields and naval transport. One of these projects which features Big Data is the '*New standard to develop Advanced Transportation Information Services*'[41]. Its aims include:

- to organize, structure and industrialize the collection of mobility data in the territories and their related actors,
- disseminate well-structured predictive and real time multimodal traveller information, properly mapped with its related infrastructure and topography,
- to promote the use of these tools for individual and collective control of the transport and mobility system at a local level.

This project uses a bottom-up, decentralized approach, based on existing diversified data issued from smart mobile applications, floating car data, institutional data and crowdsourcing. It requires Big Data storage and treatments but also highly scalable mobility analysis and visualization

2.2.13 Ireland

In 2015, Ireland launched its five-year strategy on research and development, science and technology in *Innovation 2020*[42], which sets out the roadmap for continuing progress towards making Ireland a global innovation leader, driving a strong sustainable economy and a better society. Ireland's ICT sector is world-renowned, with the world's top ten ICT companies located there. The priorities for innovation include seizing the benefits of the digital economy. Specific research priorities to underpin this vision include autonomous transport. Seven (7) priority societal challenges have been identified, where funding in research and innovation may have a real impact for Europe's citizens, one of which includes smart, green and integrated transport. INSIGHT is a Big Data and analytics research centre, involving seven national universities of Ireland. An outcome of a completed project from 2008 to 2011: '*Automated assessment of delineation and signage systems*', aimed at developing vision-based systems to evaluate road signing and lining that can be attributed to road traffic accidents when defective. The project achieved in developing a vehicular mobile mapping system composed of a number of navigation sensors for geo-referencing (positioning) and a multi-camera (stereo) system for automated feature extraction from imagery.

[40] https://strukturnifondovi.hr/
[41] https://anr.fr/Project-ANR-13-TDMO-0007
[42] https://dbei.gov.ie/en/Publications/Innovation-2020.html

2.2.14 Italy

RicercheTrasporti.it (RT)[43], is a technologies portal in Italy managed by a consultancy company, TRT Trasporti e Territorio, which collects data, information, theoretical and empirical results relating to transport policy and research, on a national and international basis. It published '*Innovation Brief on City Logistics Solutions*[44], in the context of a project (PROSPERITY), which sets out a series of policy options and innovative solutions that can bring about change to the organization of freight transport in cities. It identifies digitisation, in particular, the ability to extract value from data, especially in real time, as one of the main elements in the transport sector. The *USV Permare (*2012-2015)[45] project developed cutting edge features in the maritime sector, in particular the development of systems for monitoring coastal areas, e.gs. real-time and remote accessible on-board monitoring systems and high range/bandwidth transmission data devices.

2.2.15 Sweden

Strategic Vehicle Research and Innovation (2009-2015)[46], was a partnership between the Swedish government and the automotive industry for joint funding of research, innovation and development on climate, environment and safety. During this period, there was a collaboration program on traffic safety and automated vehicles. In a report '*Artificial intelligence in Swedish business and society – summary'*[47], AI is identified as important for Sweden's future competitiveness and innovative strength in all sectors branches. Travel and transports - autonomous vehicles, logistics and transport infrastructure and smart transports and vehicles are highlighted as particularly important development areas in AI for Sweden. '*Electric Vehicle Demonstration Programme for 2011 to 2017*'[48], was a funding program focusing on the inter-relationship between driver's needs and technical requirements. In this way, it was possible to identify opportunities and eliminate possible barriers for the large-scale introduction of electric vehicles. Several other transport studies have been carried out in Sweden and research is in progress on both innovation systems and policy measures to encourage the use of alternative motor fuels.

2.2.16 Finland

Intelligent Transport, Second Generation (2013 – 2020)[49], builds on the initial intelligent transport strategy from 2009, which is a program which supports the integration of intelligent transport into all forms of transport. ICT is used to effectively influence mobility and transport

---

[43] http://www.trt.it/
[44] http://www.trt.it/en/innovation-brief-sulle-city-logistics-solutions/
[45] http://www.insean.cnr.it/it
[46] https://www.vinnova.se/
[47] https://www.vinnova.se/en/about-us/publications-and-ebooks/?lightbox=lightbox&type=publication&url=%2Fen%2Fpublikationer%2Fartificial-intelligence-in-swedish-business-and-society%2F
[48] http://www.energimyndigheten.se/en/
[49] https://www.lvm.fi/en/intelligent_transport

options and supports development of the transport system by improving the service level of transport information. The Ministry of Transport and Communications, which is responsible for legislation and strategy development concerning information security in communications networks and services, fund projects on an ongoing basis to further its objective to build an intelligent and sustainable transport system. In terms of Big Data, the Ministry aims to increase the availability of information and open data and generate new business, promote the utilisation of automation and robotics and ensure that services and networks are safe for the users. Finland participated in many development programmes for public transport, some of which include: CAPTURE, PROGRESS, INFOPOLIS 2, MARETOPE and PROMPT. Finland also participated in several goods transport and logistics projects: EU-FRISBEE, DIFACT and EUTP II and in EU-ROSITA 2 for traffic safety.

2.2.17 Spain

'*Strategic Plan of Infrastructures and Transport (PEIT)' (2006-2020)*[50]*,* is a program which funds R&D projects in line with the Strategic Plan of Infrastructure and Transport. The project '*Analysis of accessibility and efficiency improvements of operations in modal exchange nodes and logistic platforms (INTER-NODAL)*'[51] examined issues with modal interchange nodes and logistic platforms that register daily bottlenecks, which makes the development of logistic activity difficult. It produced common methodological criteria to optimise accessibility following the analysis of actual accessibility problems and operations which affected efficiency, rapidity, productivity and problems of capacity.

2.2.18 The Netherlands

Wageningen Ede pods (WEpods) (11/14-11/15)[52] is a project that launched a pilot with two self-driving vehicles (with no steering wheel or pedals) operating on fixed routes. Additional technical equipment was equipped such as cameras, laser, radar and GPS to detect and interpret the surrounding with a view to driving safely. A control room monitors vehicle operation and safety. *Smart Mobility Schiphol (01/17-12/20)*[53], is a project that developed and tested in real life environment, communication between automated cars, traffic lights and other infrastructure. Schiphol was chosen because there are pre-existing highly intelligent traffic lights, which generates open data as well as available modalities to assess contemporaneous effects on different traffic flows. Promising groups for *Mobility-as-a-Service in the Netherlands- MaaS* (05/18-09/19)[54], is defined as a transport concept involving the use of a single digital platform to find, book and pay for trips offered by various transport service providers. Maas is a digital online platform that people will primarily access via

[50] https://www.mitma.gob.es/plan-estrategico-de-infraestructuras-y-transporte-peit
[51] https://www.itene.com/
[52] http://davi.connekt.nl/wepods-project/
[53] https://www.smartmobilityembassy.nl/smart-mobility-schiphol/
[54] https://english.kimnet.nl/publications/documents-research-publications/2019/08/15/promising-groups-for-mobility-as-a-service-in-the-netherlands

smartphones/mobile devices and faces many issues such as effective use of devices, trust in the app suppliers and technology, privacy issues etc.

2.2.19 Germany

The programme *Research Information System Mobility and Transport (FIS)*[55] was established by the Federal Ministry for Transport, Building and Housing to support the usability of public research. Ten research institutes are involved in research activities for FIS and create science-based information systems regarding mobility and transport. There are many communication papers on Big Data.

In a 2017 communication[56] various challenges facing the implementation of digital networking in freight transport and logistics are flagged, these include: increasing technical demands on the communication infrastructure, which result from the growing data traffic, the availability of sufficient frequencies in the mobile and radio network as well as for direct vehicle communication etc.

The legal framework both enables and restricts the use of mobility data. Challenges exist in respect of standards for data formats, data exchanges and process for the smooth communication between road users, vehicles and traffic infrastructure.

The increase in the amount of digital data processed in and outside motor vehicles, and related issues such as data security, GDPR rights, and autonomy/sovereignty over this type of data is flagged in a 2017 communication[57]. Vehicle data is mostly collected and processed in the vehicle, with part of it being sent to data centres, to update maps, improve data models and evaluate user data. As there is also third-party interest in this data, it compares "data as the next oil", meaning that refined data will continue to increase in value in the future. In turn, an intelligent traffic system uses high quality data sources located in vehicles and information infrastructures (e.g. INSPIRE geodata infrastructure in the EU).

2.3 Big Data challenges in the transport sector

This section discusses challenges that exist in the Big Data domain and that need to be overcome for effective usage of Big Data technologies in Transport sector.

Several research challenges related to Big Data, spread across different dimensions, have been considered under LeMO. Other projects and research (i.e., BYTE[58], BDV[59] and BPC[60]) have also considered these dimensions and challenges. The tables below provide a breakdown of these Big Data dimensions and co-related challenges.

[55] https://www.forschungsinformationssystem.de/servlet/is/1/
[56] https://www.forschungsinformationssystem.de/servlet/is/472557/?clsId0=0&clsId1=0&clsId2=0&clsId3=0
[57] https://www.forschungsinformationssystem.de/servlet/is/472557/?clsId0=0&clsId1=0&clsId2=0&clsId3=0
[58] Including BYTE project, Big Data Policy Canvas, BDVA research and innovation agenda, International Transport Forum’s research on Big Data & Transport, and European Strategy for Data.
[59] BDV, http://www.bdva.eu/
[60] Big Policy Canvas, https://www.bigpolicycanvas.eu/

Table 3 Dimension distribution of Big Data challenges in LeMO and a reference group

Standard dimension	LeMO	BYTE	BDV (Big Data Value Association)	BPC (Big Policy Canvas)	BDT (Big Data and Transport)	ESD (A European Strategy for Data)
Data management	Data resources	Data management	Data Management	Data Acquisition, Cleaning and Representativeness	Data acquisition and recording	Data interoperability and quality
	Data complexity				Data extraction, cleaning, annotation and storage	Empowering individuals to exercise their rights
					Data production: Digital vs. analogue	
					Integration, aggregation and fusion	
Data processing	Data resources	Data processing	Data Processing Architectures	Data Acquisition, Cleaning and Representativeness	Data extraction, cleaning, annotation and storage	Availability of data
	Data complexity			Data Storage, Clustering, and Integration		
Data analysis	Data resources	Data analysis	Data Analytics	Modelling and Analysis with Big Data	Analysis, modelling and visualisation	
	Data complexity			Data Storage, Clustering, and Integration	Data production: Digital vs. analogue	
					Need for transparency and metadata on data provenance	
Data visualisation		Data visualisation	Data Visualisation and User Interaction	Data Visualization	Analysis, modelling and visualisation	
Limited infrastructures and systems	Limited infrastructures and systems		Engineering and DevOps for Big Data		Data production: Digital vs. analogue	Data infrastructures and technologies
Non-technical priorities	Affecting technical solutions	Non-technical priorities	Non-Technical Aspects	Privacy, Transparency and Trust		Skills and data literacy
						Data governance
						Imbalances in market power
						Availability of data
						Empowering individuals to exercise their rights
Data protection	Affecting technical solutions	Data protection	Data Protection	Privacy, Transparency and Trust	Big Data, personal data and privacy	Cybersecurity
					Data extraction, cleaning, annotation and storage	
Big Data Standardisation	Data resources		Big Data Standardisation			

Table 4 Big Data challenges in LeMO and a reference group

Standard dimension	Big Data challenge	LeMO	BYTE	BDV (Big Data Value Association)	BPC (Big Policy Canvas)	BDT (Big Data and Transport)	ESD (A European Strategy for Data)
Data management	Unstructured data annotation	Fragmentation of data sources	Handling unstructured and semi structured data	Semantic annotation of unstructured data	Data cleaning and formatting	Heterogeneity	
	Semantic interoperability	Data interoperability	Semantic interoperability	Semantic interoperability	Real time data collection and production	Semantics of the data	Data interoperability
	Data quality assurance	Data quality	Measuring and assuring data quality	Data quality	Quality assessment	Data quality principle	Data quality
	Data lifecycle management		Data lifecycle	Data lifecycle management		Big Data collection and analysis lifecycle	
	Bias and representativeness		Bias and representativeness		Representativeness of data collected	Bias and representativeness	
	Data-as-a-Service	Data-as-a-service	Data-as-a-service model and paradigm	Data-as-a-Service			
	Data control and IPR	Data flow controls and data ownership	Data provenance, control and IPR	Data provenance			Empowering individuals to exercise their rights
Data processing	Big Data storage		Efficient mechanisms for storage	Efficient mechanisms for storage	Big Data Storage	Advances in data storage	Secure Big Data storage
	Processing heterogenous data	Data variety	Processing real-time heterogeneous data	Heterogeneity	Heterogeneity and incomplete data	Heterogeneity	
	Processing of data-in-motion and data-at-rest		Architectures for data-at-rest and data-in-motion	Processing of data-in-motion and data-at-rest			
	Decentralised architectures	Physical and logical decentralisation	Decentralised architectures	Decentralisation	Decentralized storage		Fully decentralised solutions
	Efficient algorithms for performance	Efficient data processing	Efficient mechanisms for processing	Efficient algorithms for performance		Velocity, timeliness, volatility	
	Efficient architecture selection	Efficient architecture selection		New hybrid Big Data and HPC architecture			
Data analysis	Scalable analytics	Scalability	Scalable techniques for real-time analytics	Scalability	Scalable machine learning	Scale	
	New analytic frameworks			Analytics frameworks and processing		New analytic frameworks	

	Predictive and prescriptive analytics	Predictive and prescriptive analytics	Predictive and prescriptive analytics	Predictive and prescriptive analytics			
	Event and pattern discovery		Event and pattern discovery	Event and pattern discovery	Patterns and trends identification	Causality, correlation and multiple correlations	Pattern recognition
	Collaborative model simulation		Improved models and simulations	Collaborative networks	Collaborative model simulations	Modelling	
Data visualisation	**Interactive data visualisation**		Real-time interactive visualisation	Intuitive and interactive visual interfaces	Interactive data visualization	Visualisation and dissemination	
	Data dynamic visualisation		Dynamic clustering of information	Dynamic visualisation	Dynamic real-time visualization	Dynamic visualisation	
Big Data Standardisation	**Lack of standardisation**	Lack of data standardisation	Standardisation	Technology and data standardisation	Standardisation of processes		Standardisation activity
Limited infrastructures and systems	**Limited infrastructure for real-time data processing**	Limited infrastructure for real-time processing	Improved infrastructures	Considering multiple dimensions of Big Data value	Technical infrastructure	Ubiquitous data logging and sensor platforms	Data infrastructures and technologies
Non-technical priorities	**Lack of skilled workers/expert knowledge**	Lack of skilled workers/expert knowledge	Specialist with data skills	Skills Development	Skilled and experienced staff		Skills and data literacy
	Large investment costs	Large investment costs					
	Social perceptions and societal implications	Customers' heterogeneous behaviours	Citizen research	Social Perceptions and Societal Implications			
	Open government data	Open government data	Open government data	Data governance	Open government data	Open data	Foster business-to-government data
	Discrimination discovery and prevention	Discrimination of certain societal groups	Discrimination discovery and prevention		Discrimination mining		Imbalances in market power
	Data-based new business models	New business models for data exploitation	Develop new business models	Ecosystems and Business Models		New business models for using Big Data	Data-based business models
Data protection	**Maintaining robust data privacy**	Anonymisation algorithm	Robust anonymisation algorithms	Maintaining robust data privacy	Data privacy	Traceability, privacy	Privacy preserving technologies
	Robust data protection approach	Data protection approach	Complete data protection framework	Enforceable data protection approach	Personal data protection	Personal data protection frameworks	Data protection
	Cybersecurity	Cybersecurity	Cybersecurity	Cybersecurity	Data security	Cybersecurity	Cybersecurity

A brief description of each challenge across different dimensions is provided below. Relevant technologies and techniques to overcome these challenges are discussed in Section 2.4 and their prioritization is considered in Section 2.5.

2.3.1 Data management

- ***Unstructured data annotation:*** Unstructured data such as tweet, videos and images related to transportation traffic needs to be semantically annotated in digital formats. Natural language (including multilingual text) or specific domain data must be pre-processed and enhanced with semantic annotation to extract useful and meaningful information.
- ***Semantic interoperability***: Interoperability for data storage, exchange, integration, and fusion can unlock data silos in the transport sector to allow efficient user-driven or automated annotations and transformations.
- ***Data quality assurance***: Data quality, mentioned in nearly all the LeMO case studies, assurance together with curation frameworks and workflows should include transparency on the data collection process and meta information on the context and purpose of such collection.
- ***Data lifecycle management***: With the tremendous increase in transportation data, integrated data lifecycle management is facing new challenges in handling the sheer size of data as well as enforcing consistent quality, as the data grows in volume, velocity, and variability, including supporting real-time management (Grumbach, et al. 2016).
- ***Bias and representativeness***: Although massive data sets (e.g., data from sensor networks, mobile GPSs, or geo-tagged tweets) exist, sample bias and error often hinder to correctly analyse them for discovering representative information. For example, biases may emerge from travellers' location data due to specific characteristics of app users versus the general population.
- ***Data-as-a-Service***: Data-as-a-Service is how to bundle both the data and the software and data analytics needed to interpret and process them into a single package that can be provided as exploiting new opportunities (e.g., new data-driven transportation services) for economic growth.
- ***Data control and IPR***: Data control and IPR as mentioned in LeMO case study 4 has been highlighted that certain kinds of data attract new rights and require new rights statement initiatives. Data licensing and ownership have still no means to be represented clearly for everybody, and this is especially dangerous in the Internet of things applications, where data is complex and distributed among different physical locations.

2.3.2 Data processing

- ***Big Data storage***: Big Data storage technologies are a crucial enabler for advanced analytics that has the potential to transform society and the way critical decisions are made in the transport sector. “Cloud” means remote data storage centres that allow access to and analysis of distributed data as if it were located on a single server. Not only does "cloud computing" deliver economies of scale in relation to data storage, management and support costs, it also opens up new possibilities for ad-hoc and customisable access to computing capacity on public cloud-based platforms (Crist, et al. 2015).

- ***Processing heterogenous data***: Techniques and tools for processing real-time heterogeneous data. This is particularly needed in the development of new tools for sensor data processing, especially in the transport sectors (Lyko, Nitzschke and Ngomo 2016). For instance, data fusion algorithms help process inputs from wheel movement sensors, accelerometers, magnetometers, cellular signal sensors, cameras, laser scanners and GPS chips (Crist, et al. 2015).
- ***Processing of data-in-motion and data-at-rest***: Innovative architectures for data-at-rest and data-in-motion, techniques and tools for processing real-time heterogeneous data and scalable algorithms for real-time analytics to allow optimisation of utilities through data analytics will contribute to increasing efficiency in the transport sector (Grumbach, et al. 2016). For instance, usual transportation scenarios for Big Data processing also require a greater ability to cope with systems which contain inherently dynamic in their daily operation in order to increase operational effectiveness and competitiveness.
- ***Decentralised architectures***: Big Data producers and consumers can be distributed and loosely coupled as a vehicular environment in the Internet of Things. Architectures should consider the effect of distribution on the assumptions underlying them, such as loose data agreements, missing contextual data, etc. The distribution of Big Data processing nodes requires new Big-Data-specific parallelisation techniques and the automated distribution of tasks over clusters will be a crucial element for effective stream processing (BDVA 2017).
- ***Efficient algorithms for performance***: Algorithms to reorganise complex systems and processes to improve their performance according to one or more parameters, such as travel time or fuel-efficiency has to be scaled up and reduce energy consumption compatible with the best efforts in the integration between hardware and software (Crist, et al. 2015).
- ***Efficient architecture selection***: Transportation domains have shown a huge increase in the complexity of Big Data applications, usually driven by the computation-intensive simulations, which are based on complex models and generate enormous amounts of output data. On the other hand, users need to apply advanced and highly complex analytics and processing to this data to generate insights, which usually means that data analytics needs to take place in situ, using complex workflows and in synchrony with computing platforms.

2.3.3 Data analysis

- ***Scalable data analytics***: Being able to apply storage and complex analytics techniques at scale is crucial in order to extract knowledge out of the data and develop decision-support applications (BDVA 2017).

- ***New analytic frameworks***: New frameworks and open APIs for the quality-aware distribution of batch and stream processing analytics have been required with minimal development effort from application developers and domain experts, in the transport sector (BDVA 2017).
- ***Predictive and prescriptive analytics***: Capabilities of predictive and prescriptive analytics will open up novel opportunities for predictive analytics in terms of predicting future situations, and even prescriptive analytics with regard to providing actionable insights based on forecasts, in the transport field (BDVA 2017). But even more important that such usability of the analytics results by non-data scientists, it has been recognised an urgent need of validated methodologies and standards behind the analytics on whose results decisions are to be taken (Grumbach, et al. 2016).
- ***Event and pattern discovery***: Transportation data such as travel plan generally have patterns and trends by seasons, areas and so on. Therefore, the identification of the pattern and trends is important. It includes social media analysis to identify e.g. clusters of posts that can assist to track unplanned events (like accidents) and improve emergency response. It also contains research into dynamic social processes such as the traffic jam and how Big Data can contribute to it (Grumbach, et al. 2016). Also, Big Data analytics is well suited to the discovery of correlations that were not obvious, or even visible, in the data initially. For example, combining traffic flow data with digital maps may reveal living cycle correlates to busy urban roads (Crist, et al. 2015).
- ***Collaborative model simulation***: There is a need for better integration between algorithmic and human computation approaches (Freitas and Curry 2016). Citizens should also be allowed for probing and real-time data collection for feeding model simulation at real-time, and/or contributing by mean of some sort of online platform. Understanding the present through data is often not enough and the impact of specific decisions and solutions can be correctly assessed only when projected into the future. Hence the need of simulation tools allowing for a realistic forecast of how a change in the current conditions will affect and modify the future traffic scenario (Mureddu, et al. 2019).

2.3.4 Data visualisation

- ***Interactive data visualisation***: Visualisation tools are still largely designed for the analyst and are not accessible to non-experts. Therefore, there are needs for the evolution of visual interfaces towards their becoming more intuitive and exploiting advanced discovery aspects of Big Data analytics. This is required in order to foster effective exploitation of the information and knowledge that Big Data can deliver for transport decision making (BDVA 2017).
- ***Data dynamic visualisation***: Due to continuing advances in sensor technology and the increasing availability of digital transportation infrastructure that allows for acquisition, transfer, and storage of Big Data sets, large amounts of data become available even in real-time. Since most analysis and visualisation methods focus on static data sets, adding

a dynamic component to the data source results in major challenges for both the automated and visual analysis methods (Mureddu, et al. 2019).

2.3.5 Limited infrastructures and systems

- ***Limited infrastructure for real-time data processing***: With extensive software event logging (and storage) and the deployment of millions of sensing devices enable the real-time production of petabytes of data globally in the transportation domain (Crist, et al. 2015), however, most of existing Intelligent Transport System infrastructure and traffic management solutions deployed nowadays are not sufficiently designed for real-time data processing, nor are they close to analysing the captured data at the rates demanded by critical applications (e.g. safety). Also, the design and advancement of methodologies, tooling and platforms should carefully consider the multifaceted issues of Big Data, such as real-time processing and analytics, as well as data veracity and variety (BDVA 2017). Cloud service can be one of alternatives, but there are concerns about cybersecurity and national intelligence (EuropeanComission 2020).

2.3.6 Non-technical priorities

- ***Lack of skilled workers/expert knowledge***: Currently, Big Data and analytics are top of the list of critical skills shortages. In 2017, there were approximately 496,000 unfilled positions in the area of Big Data and analytics in the EU2733. Moreover, general data literacy in the workforce and across the population is relatively low and participation gaps exist (for example by elderly people). If it is not addressed, the shortage in data experts and the lack of data literacy will affect the EU's capacity to master the challenges of the data economy and society (EuropeanComission 2020). In order to leverage the potential of Big Data Value, a key challenge for Europe is to ensure the availability of highly and appropriately skilled people with an excellent grasp of the best practices and technologies (BDVA 2017).
- ***Large investment costs***: There are large investment costs by applying Big Data technology or predictive maintenance to be used in the transport sector. Furthermore, the investment costs result from setting up and running the hardware as well as the development and management of the applications.
- ***Social perceptions and societal Implications***: Big Data will provide solutions for major societal challenges in Europe. For an accelerated adoption of Big Data, it is critical to increasing awareness of the benefits and the value that Big Data offers. End-users' lack of trust in Big Data technology is an important barrier that may hinder adoption, affecting aspects such as privacy, transparency, perceived efficacy, manageability, and acceptability. In addition, collaboration and co-innovation between organisations, the public sector and private individuals should be enhanced to support value creation from Big Data solutions (BDVA 2017).

- ***Open government data***: It is widely recognised as a method to increase trust and transparency (Domingue et al. 2016) with defining the rules for accessing and sharing data, regarding privacy and protection issues. As discussed in case study 2 of D4.2 and mentioned needs of that kind of data in many cases of LeMO and relevant projects, this issue is significantly important to also the transport sector (M. Hong, et al. 2020).
- ***Discrimination discovery and prevention***: There may be discriminations of levels for offering and utilising information between users, companies and societal groups ('Digital Divide') to use Big Data technology. It also relates to market imbalances in relation to access to and use of data, for example when it comes to access to data by SMEs. A case in point comes from large online platforms, where a small number of players may accumulate large amounts of data, gathering important insights and competitive advantages from the richness and variety of the data they hold (EuropeanComission 2020).
- ***Data-based new business models***: The Big Data Value ecosystem will comprise many new stakeholders. New concepts for data collection, processing, storing, analysing, handling, visualisation and, most importantly, usage will emerge, and business models will be created around them (BDVA 2017). Identifying sustainable business models and ecosystems in and across sectors and platforms will be an important challenge. In particular, SMEs and start-ups often require legal and regulatory advice to fully capture the many opportunities ahead from business models based on Big Data (EuropeanComission 2020).

2.3.7 Big Data standardisation

- ***Lack of standardisation***: Standards are the essential building blocks for product and service development as they define clear protocols that can be easily understood and adopted internationally (BDVA 2017). This is a prime source of compatibility and interoperability. Most technology standards for Big Data processing are de facto standards that are not prescribed by a standards organisation. However, the lack of standards is a major obstacle. Furthermore, the 'variety' of Big Data makes it very difficult to standardise. Nevertheless, there is a great deal of potential for data standardisation in the areas of data exchange and data interoperability.

2.3.8 Data protection

- ***Maintaining robust data privacy***: As discussed in most of LeMO case studies, maintaining data privacy with utility guarantees is an important challenge and one which also implies sub-challenges, such as the need for state-of-the-art data analytics to cope with encrypted or anonymised data (Hong, Akerkar, et al., Report on the characterization of the barriers and limitations 2019). Preserving anonymity often implies removing the links between data assets. However, the approach to preserving anonymity also has to be reconciled with the needs for data quality, on which link removal has a very negative impact.

- ***Robust data protection approach***: A more generic, easy to use and enforceable data protection approach suitable for large-scale commercial processing is needed. On the technical side, mechanisms are needed in order to provide data owners with the means to define the purpose of information gathering and sharing and to control the granularity at which their data will be shared with authorised third parties throughout the whole lifecycle of the data. Technical measures are also needed to enable and enforce the auditability of the principle that the data is only used for the defined purpose and nothing else (BDVA 2017).
- ***Cybersecurity***: By its nature, any data manipulation presents a cybersecurity challenge. Data Sovereignty consists in merging personal data from several sources, allowing the data owner to keep control over their own data, be it by partial anonymisation, secure protocols, smart contracts, or other methods. However, such problems have to be globally solved, taking a functionally complete and secure by design approach (BDVA 2017).

2.4 Potential Research Interventions

Building upon the sections above, the present section proposes potential research interventions that can help overcome the challenges in Big Data domain as well as promote effective usage of Big Data in Transport domain. These interventions (total 31 in number) have been divided across the 8 Big Data Dimensions discussed in Section 2.3.

2.4.1 Data management

- ***Unstructured data annotation***: The availability of homogeneous transport data from worldwide transport authorities and operators gives us the possibility of creating new types of applications related to transport (trip planners, fare calculators, ticket recommenders, etc.) that can be deployed easily in different regions or cities. Declarative mapping techniques to generate transport knowledge graphs can be used (Chaves-Fraga, et al. 2019).
- ***Semantic interoperability***: Semantic search, matching and mapping, ontology alignment, and linked data technologies should be addressed to improve interoperability between transportation data silos (Domingue, et al. 2016) (Freitas and Curry 2016). Also, contextualisation based on AI such as deep learning, machine learning, natural language and semantic analysis should be used for in near-real-time data processing (BDVA 2017)
- ***Data quality assurance***: Approaches to compute the uncertainty and algorithms to validate and annotate data need to be developed (Freitas and Curry 2016).
- ***Data lifecycle management***: Data discovery, datasets crawlers, metadata, dataset ranking (Domingue, et al. 2016), adaptive data detection, data acquisition (Lyko, Nitzschke and Ngomo 2016), and data curation could assist to data management.
- ***Bias and representativeness***: The use of Big Data by policymakers should be assessed according to the specific concerns at hand and the level of knowledge relating to a problem to be addressed.

- ***Data-as-a-Service***: Trusted and flexible infrastructures need to be developed for the DaaS paradigm, potentially based on technologies such as distributed ledgers, blockchains and/or microservices (BDVA 2017).
- ***Data control and IPR***: New theoretical models and methodologies, sandboxing and virtualisation techniques for data transportability under different contexts to grant fine-grained permission management should be developed (Freitas and Curry 2016) (Strohbach, et al. 2016).

2.4.2 Data processing

- ***Big Data storage***: To automate complex tasks and make them scalable, hybrid human-algorithmic data curation approaches have to be further developed (Freitas and Curry 2016). Basing on the advances of cloud computing, the technology market is very developed in this area, with most knowledge bases products created in the US (Sharma 2016). Crowdsourcing also plays an important role. Energy-efficient data storage methods are also a crucial research priority (Strohbach, et al. 2016).
- ***Processing heterogenous data***: Such data fusion technology is often encountered in the area of Internet of Things, and domains like smart cities. Asia possesses the whole institutions partly working on topics that are out of scope for the EU mobile operators, due to their limited subscriber bases –for example, innovative end-user mobile services (Qiao, et al. 2015).
- ***Processing of data-in-motion and data-at-rest***: Real-time analytics through event processing and stream processing, spanning inductive reasoning (machine learning), deductive reasoning (inference), high-performance computing (data centre optimisation and efficient resource allocation) and statistical analysis have to be adapted on continuous querying over transportation data streams (i.e., on-line processing) (BDVA 2017).
- ***Decentralised architectures***: An important one is efficient distribution of the processing to the Edge (i.e. local data edge processing and analytics), as a part of the ever-increasing trend for Fog computing (BDVA 2017). Also, decentralised architectures could benefit from the development of cryptographic mechanisms applicable to the cloud and Big Data, such as attribute-based encryption, which diminish the threats to data protection and personal privacy (Grumbach, et al. 2016).
- ***Efficient algorithms for performance***: Efficient algorithms should be possible to utilise existing and emerging high-performance-computing and hardware-oriented developments like main memory technology, with a different type of caches, such as Cloud and Fog computing, and software-defined storage with built-in functionality for computation near the data (e.g. Storlets[61]). Also, to be utilised are data availability guarantees to avoid unnecessary data downloading and archiving, and data reduction to support storing, sharing and efficient in-place processing of the data (BDVA 2017).

[61] https://docs.openstack.org/storlets/latest/

- ***Efficient architecture selection***: It needs to select proper existing architectures or requires novel Big Data architectures which will exploit the advantages of High Performance Computing (HPC) infrastructure and distributed processing and includes the challenges of maintaining efficient distributed data access (enabling the scaling of deep learning applications) and efficient energy consumption models in such architectures (BDVA 2017).

2.4.3 Data analysis

- ***Scalable data analytics***: The analytics based on stream data mining in contexts of a high volume of stream traffic data coming from e.g. sensor networks or large numbers of online users (Domingue, et al. 2016), as well as real-time analysis of public transportation data (Lammerant, et al. 2015) must be scalable, with low latency.
- ***New analytic frameworks***: On the other hand, many of the techniques efficiently processing Big Data have been released under open-source licenses free of commercial rights. This has greatly accelerated its uptake. For instance, map-reduce work processes (such as Hadoop or its derivatives) leverage parallel processing by breaking up large and complex semi-structured and unstructured data sets into more manageable subsets. They then allocate coordinated processing tasks to multiple distributed servers. These algorithms are fully scalable and are not bound by having to formalise database relationships ahead of storage and analysis (Crist, et al. 2015).
- ***Predictive and prescriptive analytics***: An emerging trend is to use new sensor data for predictive analysis, e.g. in Industry 4.0 (Becker, Jentzsch and Palmetshofer 2014). This field has been particularly strong in the US, showcased by such recent developments as IBM's Watson or Google's AlphaGo (Silver, et al. 2016)**.**
- ***Event and pattern discovery***: One of the most used Big Data methodologies for identification of pattern and trends is data mining. Combination of database management, statistics, and machine learning methods are useful for extracting patterns from large datasets. It also has to be taken into account that most of the Big Data is not structured and have a massive quantity of text. In this regard, text mining is another technique that can be adopted to identify trends and patterns (Mureddu, et al. 2019).
- ***Collaborative model simulation***: Modelling process should be supported by a combination of narrative scenarios, modelling rules, and e-Participation tools (all Integrated via an ICT e-Governance platform): so the policy model for a given domain can be created iteratively using the cooperation of several stakeholder groups, such as decision-makers, analysts, companies, civic society, and the general public (Mureddu, et al. 2019).

2.4.4 Data visualisation

- ***Interactive data visualisation***: For example, the immediate outcomes exhibited from the WikiCity Rome and HubCab examples (Santi, Resta and Ratti 2014) are interactive maps that invite people to engage with mobility patterns shaped by their surroundings. This could result in users electing to share a cab with a stranger to save on the cab fare or changing their mode of transport to avoid traffic congestion (Crist, et al. 2015). Also, geospatial data-based novel visualisation can benefit from the user interfaces for parallel exploration in the field of transportation.

- ***Data dynamic visualisation***: Methodologies for bringing out meaningful patterns include data mining, machine learning, and statistical methods. In the case of tools for management and automated analysis of data streams, there are CViz Cluster visualisation, IBM ILOG visualisation, Survey Visualiser, Infoscope, Sentinel Visualiser, Grapheur2.0, InstantAtlas, Miner3D, VisuMap, Drillet, Eaagle, GraphInsight, Gsharp, Tableau, Sisense, SAS Visual Analytics, Zoho Reports. In particular, Tableau[62] allows non-technical users to create interactive, real-time visualisations in minutes (Mureddu, et al. 2019).

2.4.5 Limited infrastructures and systems

- ***Limited infrastructure for real-time data processing***: The design and advancement of methodologies, tooling and platforms should carefully consider the multifaceted issues of Big Data, such as real-time processing and analytics, as well as data veracity and variety (BDVA 2017). Cloud service can be one of alternatives, but there are concerns about cybersecurity and national intelligence (EuropeanComission 2020).

2.4.6 Non-technical priorities

- ***Lack of skilled workers/expert knowledge***: New education programmes based on online curricula are needed for interdisciplinary, professional, re-skill/up-skill workforce. Such courses based on easy accessibility will stimulate life-long learning in the domain of data and the adoption of new data-related skills.
- ***Large investment costs***: LeMO Case Study found a cloud infrastructure that can be used by their customers. This could diminish the initial costs of the infrastructure. Also, some of such technologies used in the Big Data ecosystem are open source (Teoh, et al. 2019).
- ***Social Perceptions and societal Implications***: One of the interesting techniques is crowdsourcing in terms of its empowering individuals' collaboration as well as it can be used to increase data accuracy (Teoh, et al. 2019) and scale data curation (Freitas and Curry 2016), among other applications. New methods are needed to route tasks to crowdsourcing participants based on their expertise, demographic profiles, and long-term teams, and develop open platforms for voluntary work (Freitas and Curry 2016). Research is needed to better understand the social engagement mechanisms, e.g. in projects such as Wikipedia, GalaxyZoo (Fortson, et al. 2012) or FoldIt (Khatib, et al. 2011), which would amplify community engagement (Freitas and Curry 2016).
- ***Open government data***: This requires parallel actions to reduce the digital divide (Lammerant, et al. 2015). A solution is an increase of data journalists who are able to

[62] https://www.tableau.com/solutions/topic/business-dashboards

process and present such data to a wider audience (Grumbach, et al. 2016). Also, it includes the standardisation of procedures for sharing metadata, defining the (smart) contract between stakeholders, assessing technologies such as encryption and blockchain, and formulating the necessary solutions to orchestrate the agreed governance (BDVA 2017). Foster business to government data is also being required (EuropeanComission 2020).

- ***Discrimination discovery and prevention***: There are other research avenues such as data analytics, discrimination detection in data mining (Barocas 2014) (Calders and Verwer 2010) (Hajian, et al. 2012) capacity of algorithms to disadvantage users in ways exceed the legal definitions of discrimination (Sandvig, et al. 2014) (Tufekci 2015).
- ***Data-based new business models***: To develop new business models, there are suggestions as follows: open source Big Data analytics have been proposed to ensure that benefits remain in the EU (BigData 2012). However, moving beyond the Open Data Initiative to an interoperable data scheme to process data from heterogeneous sources is also seen to foster and develop new business models (BigData 2012). Other novel models are pre-competitive partnerships where organisations that are typically competitors cooperate in R&D projects of certain data value chain steps, such as data curation, that do not affect their competitive advantage and public-private partnerships (Freitas and Curry 2016).

2.4.7 Big Data standardisation

- ***Lack of standardisation***: One example of technologies is the NoSQL databases. The history of NoSQL is based on solving specific technology challenges that lead to a range of different storage technologies (BDVA 2017). For data standardization, enabling the seamless flow of data between participants (i.e. companies, institutions and individuals) is a necessary cornerstone of the ecosystem (BDVA 2017). For example, DATEX for traffic and travel information centres is a European standard for the exchange of traffic information (Teoh, et al. 2019).

2.4.8 Data protection

- ***Maintaining robust data privacy***: Anonymisation schemes may expose weaknesses exploitable by opportunistic or malicious opponents, and thus new and more robust techniques ensuring the irreversibility must be developed to tackle these adversarial models. On the other hand, encrypted data processing techniques, such as multiparty computation or homomorphic encryption, provide stronger privacy guarantees but can currently only be applied to small parts of a computation due to their great performance penalty. Furthermore, measures to quantify privacy loss and data utility can be used to allow end-users to make informed decisions.
- ***Robust data protection approach***: In terms of data store, the major challenges are now in non-relational data stores (Strohbach, et al. 2016). This includes a better scalable

transaction model in data protection law (Lammerant, et al. 2015). Although distributed trust technologies such as blockchains can be part of the solution in distributed settings such as supply chains (BDVA 2017), considerable research is required to better understand how data can be misused, how it needs to be protected and integrated (Strohbach, et al. 2016).

- ***Cybersecurity***: In terms of research and innovation, a number of topics have to be considered as follows: homomorphic encryption; threat intelligence and how to test a learning process; assurance in gaining trust; differential privacy techniques for privacy-aware Big Data analytics; the protection of data algorithms. It is also important to define the concepts of measurable trust and evidence-based trust. Data should be secured at rest and in motion.

2.5 Prioritization of research interventions

This section explores the linkages between the *priorities* (identified through STRIA in Section 2.1 and across different countries in Section 2.2) and the *Big Data challenges* (across different Dimensions mentioned in Section 2.3).

Being the first draft of the research roadmap, the aim here is to present very preliminary findings, paving way for final recommendations. The insights gained here will be further refined by integrating findings from the survey conducted within the framework of the policy roadmap, answers received from the audience during the 2nd LeMO Webinar polls, as well as discussion with the LeMO advisory and reference board during the final validation workshop.

Ultimately, the refined outcomes will be used to prioritize the research interventions discussed in Section 2.4 above and develop final recommendations.

2.5.1 Linkages between Big Data dimensions and STRIA roadmap priorities

The table below indicates Big Data dimensions & challenges and their links with STRIA roadmap priorities. Out of the 7 STRIA areas, Big Data finds a prominent reference in 5 transport areas (TA) namely: *Cooperative, Connected and Automated Transport (TA1)*; *Vehicle Design and Manufacturing (TA2)*; *Network and Traffic Management Systems (TA3)*; *Smart Mobility and Services (TA4)*; and *Transport Infrastructure areas (TA5)*. Accordingly, these 5 areas have been chosen for the purpose of identifying linkages.

Table 5 Relation of BD challenge with Strategic Transport Research and Innovation Agenda

Big Data dimension	Challenges	TA1	TA2	TA3	TA4	TA5
Data management	Unstructured data annotation					
	Semantic interoperability					
	Data quality assurance					
	Data lifecycle management					

	Bias and representativeness					
	Data-as-a-Service					
	Data control and IPR					
Data processing	Big Data storage					
	Processing heterogenous data					
	Processing of data-in-motion and data-at-rest					
	Decentralised architectures					
	Efficient algorithms for performance					
	Efficient architecture selection					
Data analysis	Scalable analytics					
	New analytic frameworks					
	Predictive and prescriptive analytics					
	Event and pattern discovery					
	Collaborative model simulation					
Data visualisation	Interactive data visualisation					
	Data dynamic visualisation					
Big Data Standardisation	Lack of standardisation					
Limited infrastructures and systems	Limited infrastructure for real-time data processing					
Non-technical priorities	Lack of skilled workers/expert knowledge					
	Large investment costs					
	Social perceptions and societal implications					
	Open government data					
	Discrimination discovery and prevention					
	Data-based new business models					
Data protection	Maintaining robust data privacy					
	Robust data protection approach					
	Cybersecurity					

It can be seen from above that challenges relating to 'Semantic interoperability', 'Data quality assurance', 'Predictive and prescriptive analytics' and 'Event and pattern discovery' have found a reference across all STRIA roadmaps. While challenges such as 'Lack of skilled workers/expert knowledge' and 'Data-based new business models' have not found a reference across any of the transport areas under STRIA.

2.5.2 Linkages between Big Data dimensions and programs/strategies/initiatives in EU and other countries

The graph below indicates Big Data dimensions and their appearance in the programs/strategies/initiatives of the EU and other countries mentioned in Section 2.2. The analysis revealed that out of the 18 countries only 15 discussed challenges across different data dimensions. Accordingly, 3 countries (namely India, UK and Croatia) do not find a mention in the analysis below.

Though this is not an exhaustive representation of the programs/strategies/initiatives at the EU or country level, still key documents have been consulted and an attempt at co-relation with the Big Data dimensions has been made.

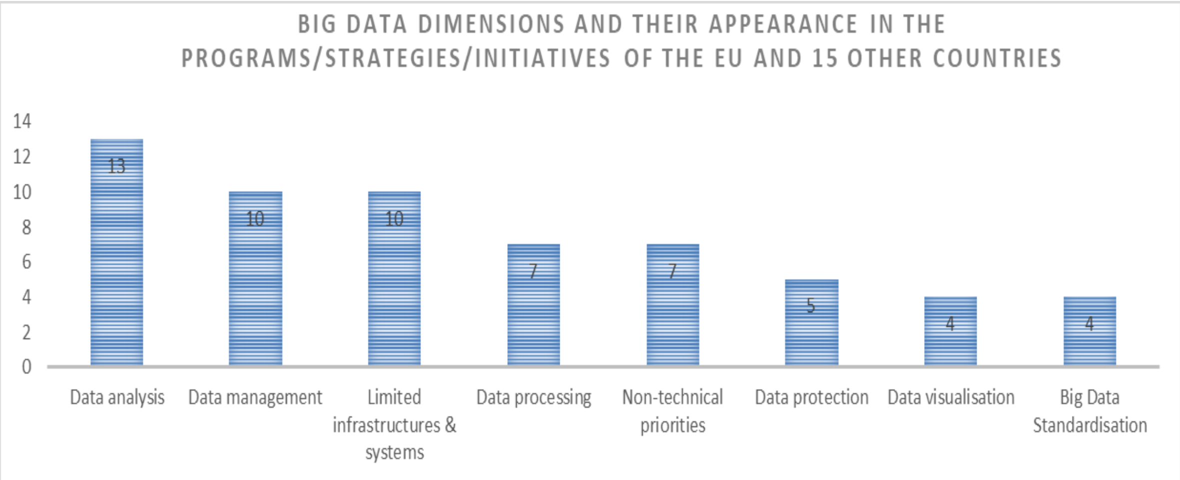


Figure 3 Big Data dimensions considered in the EU and 15 other countries

From the graph, it emerges that challenges across dimensions 'Data Analysis', 'Data Management' and 'limited infrastructure & systems' find a reference across a greater number of countries. While challenges across 'data visualization' and 'standardization' find lesser number of references.

2.5.3 Solution needs for BD challenges according to the European strategy of data

In light of the European Data Strategy, the table below presents a timeline for deploying different solutions to address the challenges across different Big Data dimensions.

Table 6 Big Data challenges timeline related to 'European strategy of data'

	Transport Research Areas in the context of Big Data	Q3 2020	Q4 2020	2021	2022	2025	2027
Data management	Unstructured data annotation			■			
	Semantic interoperability	■	■	■	■	■	■
	Data quality assurance			■	■	■	■
	Data lifecycle management						
	Bias and representativeness						
	Data-as-a-Service		■	■	■		
	Data control and IPR			■	■		
Data processing	Big Data storage	■	■	■	■	■	■
	Processing heterogenous data			■			
	Processing of data-in-motion and data-at-rest			■			
	Decentralised architectures			■	■	■	■
	Efficient algorithms for performance			■	■	■	■
	Efficient architecture selection				■	■	■
Data analysis	Scalable analytics			■			
	New analytic frameworks				■	■	■
	Predictive and prescriptive analytics						
	Event and pattern discovery						
	Collaborative model simulation			■	■	■	■
Data visualisation	Interactive data visualisation			■	■		
	Data dynamic visualisation			■	■		
Big Data Standardisation	Lack of standardisation			■	■	■	■
Limited infrastructures and systems	Limited infrastructure for real-time data processing			■	■	■	■
Non-technical priorities	Lack of skilled workers/expert knowledge	■	■	■	■	■	
	Large investment costs			■	■	■	■
	Social perceptions and societal implications			■	■	■	■

	Open government data						
	Discrimination discovery and prevention						
	Data-based new business models						
Data protection	Maintaining robust data privacy						
	Robust data protection approach						
	Cybersecurity						

As mentioned previously, the research and analysis presented in this first version of the roadmap need to be further enriched and validated. This will be done within the framework of Task 4.5 of the project. In view of this, the outcomes above should be seen as a step towards the development of final research recommendations under the LeMO project. It would be updated and finalized under Task 4.5.

3 Policy Roadmap

Section 3 presents the first version of the policy roadmap under the LeMO project. At this stage, the focus of the work is the prioritization of challenges to be met by the policy roadmap and the long list of policy recommendations that will constitute the final roadmap.

The comprehensive work of identifying the main issues have been presented in D4.1, resulting in a set of 129 different issues. It is the aim of the work here to reduce the list of issues, by identifying the priority and urgency of the issues according to experts and stakeholders. The process of identifying the main issues to be dealt with will be presented in Section 3.1.

To address these issues, a long list of recommendations is presented furnished. This is based on the work conducted in D4.2, which identified 114 different opportunities and recommendations. The list of recommendations will serve as a basis for the workshops that will be conducted to furnish the final policy roadmap in Task 4.5 of the LeMO project.

3.1 Key challenges to be addressed

The procedure used to identify the key challenges to be addressed in the policy roadmap is based on expert and stakeholder input. To gather this input, a survey was prepared and sent out to over 400 organizations and experts, mostly located within the EU or the EEA.

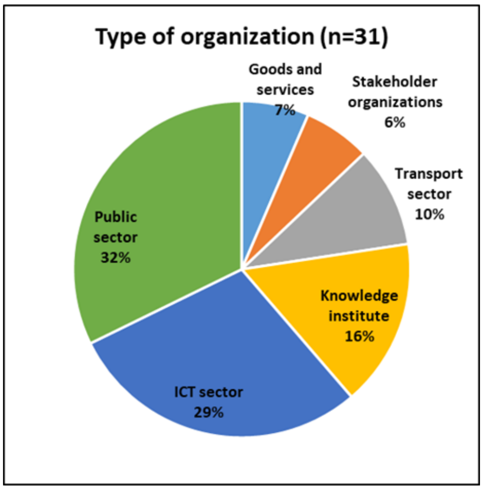
The aim of the survey was to gather their opinion on the issues that have been identified by the consortium and presented in D4.1. A pragmatic step in the procedure is the reduction of the 129 different issues by bundling issues together or by elimination. The consortium worked on reducing the list accordingly, with the surveyed issues presented in Appendix A. The list resulted in 22 unique, concise and condensed issues that was used in the survey. The questionnaire used is presented in Appendix B. The survey was administered using the EU Survey tool at the url[63].

At the writing of the report, only 31 responded to the survey. The survey will remain open until next deliverable is to be submitted. The representativeness of the results is presented in Figure 4(a-d).

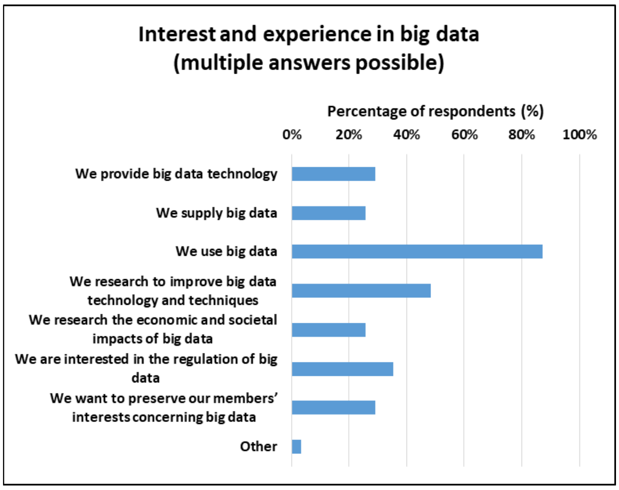
The survey asked respondents to rate the extent to which each of the 22 issues would be limiting to the further application of Big Data in the transport sector. The possible range of answers: “Not limiting, Slightly limiting, Moderately limiting, Very limiting, Extremely limiting, and No comment, insufficient information to judge”. The first five options are ordinal answers, and the last gives the respondent the opportunity to refrain from answering. In addition, the survey provided free text space for respondents to give an extended opinion or qualification of their answer.

Each of the 22 issues were scored using the ordinal answers: the answers “Not limiting” to “Extremely limiting” were valued from 1 to 5, respectively. The score for each issue was calculated as the average score given. 14 issues of the 22 issues surveyed were found to have a score of at least 3.0 “Moderately limiting”, which is the threshold applied for the rest of the roadmap.

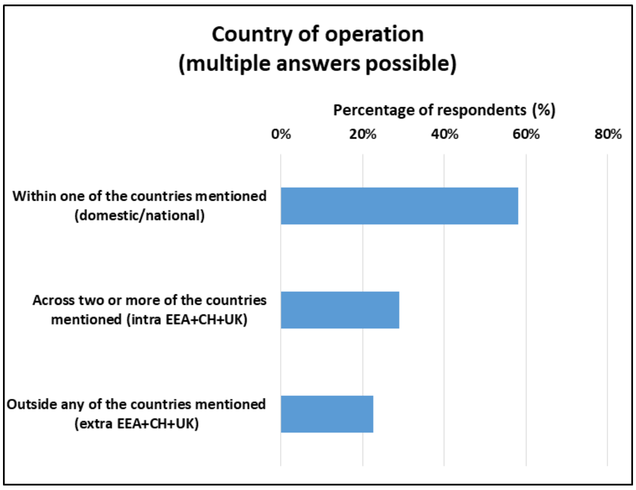
[63] https://ec.europa.eu/eusurvey/runner/BigDataIssuesInTransport2020



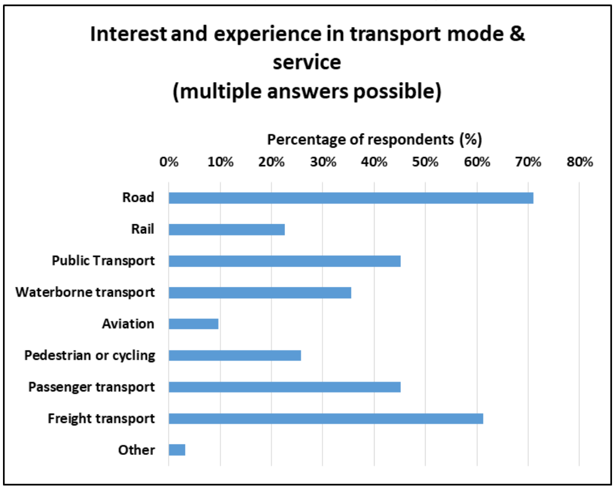
(a) Type of organization



(b) Interest and experience in Big Data



(c) Country of operation



(d) Interest and experience in transport mode & service

Figure 4 Sample of the survey

Table 7 Ranked list of issues in survey with average score above 3.0 with number of ordinal answers given

Title	Issue description	Average score	Number of answers
4. Competition	Refusals to share data may in certain cases raise abuse of dominant position concerns, especially as regards to leading technology companies (such as Google, Amazon, Facebook, Apple and Microsoft).	3.6	28
6. Privacy and data protection	Compliance with the General Data Protection Regulation (GDPR) raises technical challenges regarding the principles of data minimization, purpose limitation and transparency, relying on a data subject's consent, and responding to requests from data subjects.	3.4	31
1. Collecting data	Gathering the right data to develop transport-related Big Data technologies remains virtually impossible. Transport data is often simply inaccessible, expensive to procure or collect, spread across multiple sources, or not available in digital form.	3.3	30
13. Data ownership	Despite the absence of specific ownership right in data, multiple actors involved in the Big Data value chain could try to claim ownership in (parts of) a dataset, which would hinder the production of, access to, linking and re-use of Big Data, including in the transport sector.	3.3	29

15. Liability	The current unclear, non-harmonised EU legal framework on liability entails legal uncertainty and accordingly stifles the uptake of Big Data in the transport sector.	3.3	23
8. Mixed datasets	Big Datasets commonly used in the transport sector are often composed of both non-personal data and personal data that are inextricably linked. The comprehensive GDPR will apply fully and without limitation to all data contained in such datasets, thus severely reducing the value of it.	3.2	28
9. Data quality	There is a high risk that the quality of transport data is low and unreliable. Extracting value in light of poor data quality increases the costs of processing, risks of error and failure, and hampers the growth of the data industry.	3.2	30
14. Data sharing agreements	Contractual agreements for data sharing cannot be enforced vis-à-vis third parties. This entails that no recourse is available against third parties that obtain unjustified access to or misuse the data.	3.2	27
18. Data professionals	There is a shortage of data scientists, engineers, analysts, and developers in the local job market.	3.2	30
19. Data standardization	There are a variety of transport data formats, standards and technologies used within the transport sector, even by organizations providing the same services or by Member States across the EU. The lack of data technology standardisation reduces collaborative research, large-scale analytics and the sharing of sophisticated tools and methodologies.	3.2	29
21. Change management	Actors in the transport sector are slow and unwilling to adapt their organization to incorporate the use of Big Data technology. Some issues are the unwillingness to experiment, to restructure the organization and business practices, and to realign and update the different internal data policies.	3.2	28
5. Surveillance	Surveillance, by public or private actors, enabled by aggregated Big Data analysis, threatens individual privacy. The threat of surveillance is likely to lead to further restrictions on the use of Big Data technology.	3.1	29
20. Open data licenses	Although standard open data licences are encouraged by the PSI Directive (Directive 2003/98/EC on the re-use of public sector information), licences are still widely diverging in different EU Member States. Any company that wishes to reuse PSI from different Member States with the aim of developing a product is obliged to take into account a multitude of licences.	3.1	23
17. Data security	Protecting data from unauthorized use is becoming more complex and expensive. Data security, which involves data encryption, tokenization, and key management practices that protect data across all applications and platforms, is a highly specialized function that few organizations are equipped to handle.	3.0	29

3.2 Long list of recommendations

In the previous deliverable D4.2, the consortium provided 114 opportunities and interventions to deal with the issues identified in D4.1. The short list of issues identified via the survey will be used to also create a long list of recommendations and interventions that will be used to build the policy roadmap. The list of recommendations is presented below.

3.2.1 Competition

Competition between industry actors can be strongly reduced, if certain actors refuse to share data, which may be needed in the business models of others. Some of the recommendations to address the issue:

- Define legally the infringement of competition laws in light of the acquisition of data-rich companies under merger control regimes. The key criteria should be the kind of data to be acquired, how unique it is, whether it can be easily replicated, and whether rivals can be shut out. (OPT-LEG-COM-1)

- Support the collection of new types of data, to increase the variety of data source options. (OPT-LEG-COM-2, -3)

3.2.2 Privacy and data protection

The protection of privacy and personal data should be balanced with the ability for industry actors to extract value from the data collected, as well as to pursue new ways to provide services. The development of policy could aim and take advantage of the following measures.

- The development of coordinated and EU-wide guidance and templates, taking into account complex data processing activities, at EU and national level will likely increase legal certainty for those involved in the data value chain, and ultimately benefit data subjects. (OPT-LEG-DP-2)
- New consent models (to give consent for the use of data) should be developed in line with the operational aspects of how Big Data is used, i.e. often with automation. This should cover both the collection and the withdrawal of consent. (OPT-LEG-DP-4)
- Actors of the data value chain, including authorities, standardisation bodies, service providers, vendors and industry players can develop together standards, certification mechanisms, seals, marks and codes of conduct, which can improve accountability and compliance with the GDPR. (OPT-LEG-DP-9)
- Anonymisation techniques may serve as a means to comply with data protection law, and specifically with the following obligations: (i) data protection by design and by default; (ii) security of processing; (iii) purpose limitation; and (iv) storage limitation. (OPT-LEG-AP-3)
- Promote technical solutions to reduce the privacy problem due to data including personal information. For example, classifying personal level of data, protecting data by encryption and adopting data sensitivity for trade-off between privacy and utilising data. (OPT-ES-PRI-1)

3.2.3 Collecting data

The challenge with collecting data stems from a variety of factors, including the issues of competition and data protection. The focus here is on technical recommendations, which may not be new, but worth pursuing further.

- Existing data platforms to sell or share data should be further replicated or extended into different domains and geographical scope. There are a variety of attempts to set up Industrial Data Platforms, Personal Data Platforms (IDPs/PDPs), Research Data Platforms and Urban/City Data Platforms as marketplaces for providing and sharing data targeting specific audiences and providing specific types of data. Some examples are European data portal[64], Mobilitäts Daten Marktplatz[65], or GoData[66]. While it is unclear whether these

[64] https://www.europeandataportal.eu/data/
[65] https://www.mdm-portal.de/about-mdm/technical-details/?lang=en
[66] https://www.govdata.de/

data platforms are effective, the basic idea behind the business model certainly sounds effective. Further research is needed to clarify the success factors. (OPT-TM-DR-1)

- Support for disseminating data collection and processing methods and technology could be increased. The Open Source community is an important means to the development, experimentation and adoption of newer and better methods. Key methods include API network management and cloud platforms used to facilitate data collection, stream processing, machine learning and data acquisition. (OPT-TM-DC-2)

3.2.4 Data ownership

Multiple claims of ownership for a particular dataset (or parts of it) could hinder the data economy. Some recommendations to be explored include:

- Make clarification of ownership via adopting advanced techniques such as blockchain (OPT-ES-DO-1)
- To avoid threats for user rights of personal data ownership, it should be strictly regulated to ask users' agreement and to use the data within the necessary processes, since it is also related to the privacy issue (OPT-ES-DO-2)

3.2.5 Liability

The uncertainty in the current EU framework on liability stifles the uptake of Big Data in the transport sector.

- Anonymisation techniques may prove to be apt instruments to protect non-personal information in a technical manner. If successful, this may encourage stakeholders involved in the Big Data value cycle to engage in data sharing. (OPT-LEG-AP-4)
- If stakeholders in the Big Data analytics lifecycle are able to rely on intellectual property rights to protect their investment (be it monetary or intellectual) in (parts of) the datasets, they may be more eager to engage in data sharing in a Big Data analytics context, including in the transport sector. (OPT-LEG-IP-1)
- The ongoing work of the EU institutions regarding extra-contractual and statutory liability specifically, and the envisaged non-regulatory and regulatory interventions, may be beneficial for the uptake of Big Data in the transport sector and improve legal certainty in relation thereto. (OPT-LEG-LI-1)

3.2.6 Mixed datasets

Mobility data may often contain personal data, which would render the full database subject to the tight restrictions placed on personal data.

- So as not to require the separation of personal data from the mixed dataset (e.g. personal data contained in data sets on vehicles), policy should permit only select authorised actors to access and use the data. (OPT-ES-PRI-2)

3.2.7 Data quality

There is a need to support the quality assurance for data produced or shared.

- There is a general need for the creation of data curation frameworks and workflows. This includes i.e. data curation pipelines, on-line and off-line data filtering techniques, etc. (OPT-TM-DR-2)
- Guidance and funding for improving data sources. For instance, this might involve fixing or replacing the data collection systems, such as ITS systems, vehicle sensors, and traffic cameras. Older systems could be replaced with newer ones that produce better data quality, more suitable for BDA, i.e. 5Vs. If sources are external to the company, contractual best practices could be considered to ensure that data quality is enforced as part of the contractual best practices. Further research might be necessary to understand how best to monitor data quality. (OPT-EP-DA-1)
- Research to develop new techniques to augmenting existing poor-quality data sources. i.e. modify how data analytics is performed. Multiple data sources could be used, which increases the robustness of the analysis and reduces reliance on the lower quality data source. Augmented ML might also be researched to supplement machine learning with human input. (OPT-EP-DA-2)

3.2.8 Data sharing agreements

Contractual agreements for data sharing cannot be enforced vis-à-vis third parties, thus reducing options for legal recourse against third parties that obtain unjustified access to data or the misuse of it.

- Currently, data sharing agreements provide the only solution to govern access to and/or exchange of data between the numerous stakeholders active in the Big Data value cycle, including in the transport sector. The contractual freedom left to the parties when drafting data sharing and related agreements allows them to shape their arrangements as they see fit on the basis of the many possible varying circumstances. Governments could provide support, guidance and templated for ensuring fair data sharing agreements, especially for smaller firms/actors that do not have adequate legal resources. (OPT-LEG-DSA-1, 2)

3.2.9 Data professionals

The shortage of professionals in the labour market is an issue that may solve itself eventually, but it could use support from the government.

- Support the use of remote working and short-term employment. Gig economy platforms can help to fill the gap in personnel in the very short term, but new methods of securely, effectively and creatively working together must be developed and researched. (OPT-EP-OC-1)
- Immigration policy to be simplified and modified to attract foreign talent (or migrated citizens) from outside the EU. Different visas with different incentives could be proposed

to ensure that those with specific skillsets or have started businesses needed to boost the sector and investment in the EU. (OPT-EP-OC-2)

- Develop training and education programs ready for the future skills demand. Industries and universities should work together to develop relevant industry driven training programs. Research funding made available to attract researchers to academia and to develop data science research clusters. (OPT-EP-OC-3)

3.2.10 Data standardization

Standardization always remains a hotly debated topic. In some use cases, standardization will help the industry move forward, but this should be closely discussed with as many stakeholders from the industry as possible, so as not to stifle instead entrepreneurship.

- Creating value via data exchange will attract companies towards the standards needed for cooperation. For instance, the APIs used by public transport/infrastructure authorities, open-source solutions, and data platforms. (OPT-TM-DR-3, OPT-TM-DC-2)
- The current policy framework is characterized by a drive towards harmonization of various legislative frameworks across the EU Member States. This has created a better level playing field for companies wishing to deploy cross-border data-driven applications and technologies. (OPT-POL-PU-4)
- The adoption of soft law in relation to data, such as in particular in the context of data sharing, has positively triggered a change in mentality among private actors on the market. (OPT-POL-PU-6)
- Public tenders could be an alternative means of imposing data sharing obligations and data standards. (OPT-LEG-DS-4)
- Adopting standards and best practices for data sharing with the right access controls. Encouraging the industry to adopt standards happens gradually if the standards have concrete value. Research to develop the right kind of standards still needed. (OPT-EP-DA-3)

3.2.11 Change management

Organizations (private and public) may need support to effectively steer their organization towards a successful adoption of Big Data technology. Governments may support their efforts.

- Support change management. Organizations need help to coordinate the organizational structure and processes to effectively and efficiently make use of BDA driven insights. Public authorities and third parties could support change management via three approaches: Guidelines, Tools and methods, and financial incentives. (OPT-EP-OC-4)
- Encourage start-ups and spin-offs. Organizations that find it costly and difficult to reinvent their organization may consider instead supporting spin offs or start-up like ventures. If successful, the organization could be reintegrated into the company or allowed to operate independently. Funding to support start-ups could also be made available from public funds, which support entrepreneurship and social causes. (OPT-EP-OC-5)

3.2.12 Surveillance

Surveillance, by public or private actors, enabled by aggregated Big Data analysis, may pose a threat to individual privacy, which will hinder and oppose the growth of Big Data adoption, where personal data is concerned.

- Guidance on how to reduce (unnecessary) surveillance could be provided. For instance, how to avoid installation of sensors/cameras in places, where any personal information could be included, and strictly regulating access permissions according to levels of surveillance data to reduce misuses could alleviate serious concerns of the surveillance issue. (OPT-ES-SUR-1)
- A threshold system where data for cities below a certain activity level threshold is not made public could be adopted to reduce the threats about identifying a specific person via Big Data analysis. (OPT-ES-SUR-2)

3.2.13 Open data licenses

Harmonization of licenses and adoption of open data still needs further development.

- Data sharing obligations imposed through legislation may offer opportunities for increased competition and innovation by opening up data to private actors which would otherwise not have access to such data. (OPT-LEG-DS-1)
- Typically, when data must be shared with other private actors, we see that some kind of remuneration may be demanded, allowing the businesses involved to recover the related costs. This may be considered as alternative to free open data. (OPT-LEG-DS-2)

3.2.14 Data security

Ensuring security of data is a prerequisite to any business venture, especially those that deal with sensitive information (whether commercial or personal).

- Actors of the data value chain, including authorities, standardisation bodies, service providers, vendors and industry players can develop together standards, certification mechanisms, seals, marks and codes of conduct, which can improve compliance with various legislations such as the GDPR or the NIS Directive. (OPT-LEG-CBO-3)

The first version of the policy roadmap as presented in this chapter will be further elaborated under task 4.5. First, the survey remains open until the deadline of the next deliverable. Should there be more response to the survey this will be added to the analysis already done. Secondly, the long list of recommendations resulting from this analysis shall be discussed with stakeholders in a workshop setting. During this workshop, key recommendations for the policy roadmap will be identified. Finally, output of the workshop shall be used to provide the final policy recommendations under the LeMO project. This work will be carried out under Task 4.5.

4 Conclusion

This document is a step towards the development of a research and policy roadmap to foster the use of Big Data in the transport sector and promote sustainable transportation.

In this first version, the longlist of challenges/issues identified in previous work done under the LeMO project have been organised and consolidated to prioritise the challenges/issues and develop the final recommendations based on such priorities.

Under the research roadmap, the work done in previous LeMO deliverables has been expanded by additional analysis of relevant research issues identified in other roadmaps, EU data strategies, H2020 programmes as well as programs/strategies/initiatives of 18 countries both within and outside the EU. This has resulted in a list of 31 research issues across 8 Big Data domains namely *Data Management*, *Data Processing*, *Data Analysis*, *Data Visualisation*, *Limited Infrastructure and Systems*, *Non-technical priorities*, *Big Data Standardization and Data Protection*. For each issue, a first inventory of potential research interventions is also presented.

For the policy roadmap, a list of 22 issues have been identified. This short list of issues has been used in a survey, in which experts are asked to provide their opinion on the extent to which each of these issue would be limiting to the further application of Big Data in the transport sector. An initial analysis of the results points to that *competition concerns, challenges regarding privacy and data protection, difficulties in data collection, uncertainties surrounding rights data ownership and liability concerns* are among the issues experienced by the experts as most limiting.

In the next step towards the development of research and policy roadmap, the research and analysis presented in this document shall be further enriched and validated. This work will be done under Task 4.5 of the LeMO project. Part of the work foreseen under this task is a validation workshop with a group of experts. The aim of this workshop is to come to a final prioritised list of interventions and recommendations for both the roadmaps.

In view of this, the outcomes in this document should be seen as preliminary and not as conclusive or final.

References

Barocas, Solon. 2014. *Data mining and the discourse on discrimination.*

BDVA. 2017. *European big data value strategic research and innovation agenda.* Big Data Value Association, Tech. Rep.

Becker, Tilman, Anja Jentzsch, and Walter Palmetshofer. 2014. *Cross-sectorial roadmap consolidation.* BIG Project.

BigData. 2012. *Big Data: A New World of Opportunities.* NESSI White Paper.

Calders, Toon, and Sicco Verwer. 2010. "Three naive Bayes approaches for discrimination-free classification." *Data Mining and Knowledge Discovery* (Springer) 21 (2): 277-292.

Chaves-Fraga, David, Adolfo Anton, Jhon Toledo, and Oscar Corcho. 2019. *ONETT: Systematic Knowledge Graph Generation for National Access Points.*

Crist, P., E. Greer, C. Ratti, P. Humanes, G. Konzett, J. Tijink, D. Figuero, and R. Lax. 2015. *Big Data and Transport: Understanding and assessing options.* International Transport Forum Data Base.

Domingue, John, Nelia Lasierra, Anna Fensel, van Kasteren Tim, Martin Strohbach, and Andreas Thalhammer. 2016. "New Horizons for a Data-Driven Economy: Big Data Analysis." In *New Horizons for a Data-Driven Economy*, by José María Cavanillas, Edward Curry and Wolfgang Wahlster, 63-86. Springer, Cham.

EuropeanComission. 2020. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee And the Committee of the Regions--Europe 2020 Flagship Initiative Innovation Union.* Bruxelas: EU Commission.

Fortson, Lucy, Karen Masters, Robert Nichol, EM Edmondson, C. Lintott, J. Raddick, and J. Wallin. 2012. "Galaxy zoo." *Advances in machine learning and data mining for astronomy* 213-236.

Freitas, Andre, and Edward Curry. 2016. *Big data curation.* Springer, Cham.

Grumbach, Stephane, Aurelien Faravelon, Marti Cuquet, Anna Fensel, Scott Cunningham, and Rachel Finn. 2016. *BYTE Policy and Research Roadmap.*

Hajian, Sara, Anna Monreale, Dino Pedreschi, Josep Domingo-Ferrer, and Fosca Giannotti. 2012. "Injecting discrimination and privacy awareness into pattern discovery." *2012 IEEE 12th International Conference on Data Mining Workshops.* IEEE. 360-369.

Hong, Minsung, Rajendra Akerkar, Tharsis Teoh, Timo Eichhorn, and Julien Debussche. 2019. *Report on the characterization of the barriers and limitations.* LeMO.

Hong, MinSung, Rajendra Akerkar, Tharsis Teoh, Timo Eichhorn, Jayant Sangwan, and Julien Debussche. 2020. *Horizontal Analysis and Socio-Economic Impacts.* LeMO.

Katrakazas, Christos, Constantinos Antoniou, Vazquez, Natalia Sobrino, Ilias Trochidis, and Stratos Arampatzis. 2019. "Big Data and Emerging Transportation Challenges: Findings from the NOESIS project." *6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS).* Cracow, Poland. 1-9.

Khatib, Firas, Frank DiMaio, Seth Cooper, Maciej Kazmierczyk, Miroslaw Gilski, Szymon Krzywda, Helena Zabranska, et al. 2011. “Crystal structure of a monomeric retroviral protease solved by protein folding game players.” *Nature structural & molecular biology* (Nature Publishing Group) 18 (10): 1175-1177.

Lammerant, Hans, Paul De Hert, Guillermo Vega Gorgojo, and Erik Stensrud. 2015. *Evaluating and addressing positive and negative societal externalities.* BYTE.

Lyko, Klaus, Marcus Nitzschke, and Axel-Cyrille Ngonga Ngomo. 2016. *Big data acquisition in New Horizons for a Data-Driven Economy.* Springer, Cham.

Mureddu, Francesco, David Osimo, Esther Garrido, Ricard Munné, Juliane Schmeling, Vittorio Loreto, Peter Parycek, Gianluca Misuraca, and Giuseppe Veltri. 2019. *Roadmap for Future Research Directions.* BigPolicyCanvas.

Neilson, Alex, Indratmo, Ben Daniel, and Stevanus Tjandra. 2019. “Systematic Review of the Literature on Big Data in the Transportation Domain: Concepts and Applications, Big Data Research.” *Big Data Research* 35-44.

Qiao, Xiuquan, Sida Xue, Junliang Chen, and Anna Fensel. 2015. “A lightweight convergent personal mobile service delivery approach based on phone book.” *International Journal of Communication Systems* (Wiley Online Library) 28 (1): 49-70.

Robert, P., Biuk-Aghai, T., Kou Weng, and Fong Simon. 2016. “Big data analytics for transportation: Problems and prospects for its application in China.” *IEEE Region 10 Symposium (TENSYMP).* Bali, Indonesia: IEEE. 173-178.

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. “Auditing algorithms: Research methods for detecting discrimination on internet platforms.” *Data and discrimination: converting critical concerns into productive inquiry* 22.

Santi, Paolo, Giovanni Resta, and Carlo Ratti. 2014. “Quantifying the Benefits of taxi trips in New York through Shareability Networks.” Edited by Citeseer. *Smart Cities* 25.

Sharma, Sugam. 2016. “Expanded cloud plumes hiding Big Data ecosystem.” *Future Generation Computer Systems* (Elsevier) 59: 63-92.

Silver, David, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, et al. 2016. “astering the game of Go with deep neural networks and tree search.” *nature* (Nature Publishing Group) 529 (7587): 484.

Strohbach, Martin, Jorg Daubert, Herman Ravkin, and Mario Lischka. 2016. *Big data storage in New horizons for a data-driven economy.* Springer, Cham.

Teoh, Tharsis, Bastiaan van Berne, Ivo Hindriks, Rut Waldenfels, Timo Eichhorn, Todor Ivanov, Minsung Hong, et al. 2019. *Case study reports on constructive findings on the prerequisites of successful big data implementation in the transport sector.* LeMO.

Tufekci, Zeynep. 2015. “Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency.” *Colorado Technology Law Journal* (HeinOnline) 13: 203.

Welch, Timothy F., and Alyas Widita. 2019. “Big data in public transportation: a review of sources and methods.” *Transport Reviews* 795-818.

Zannat, Khatun E., and Charisma F. Choudhury. 2019. "Emerging Big Data Sources for Public Transport Planning: A Systematic Review on Current State of Art and Future Research Directions." *Journal of the Indian Institute of Science* 601-619.

Appendix A: Bundling of limitations

D4.3 Issue	Number of issues	Limitations (in code)
1. Collecting data. Gathering the right data to develop transport-related Big Data technologies remains virtually impossible. Transport data is often simply inaccessible, expensive to procure or collect, spread across multiple sources, or not available in digital form.	3	LIM-TM-DR-1; LIM-TM-DC-2; LIM-TM-LI-1
2. Availability of open data. Too few transport-related datasets are available as open data. While some are simply not published, although required to, some remain inaccessible due to bureaucracy, technical incompatibility, or unnecessarily restrictive licensing.	1	LIM-TM-DR-3
3. Intellectual property rights. The recent EU copyright reform introduced exceptions (covering both copyright and database right) to the exclusive rights of right holders. One of the exceptions will likely lead to text and data miners exploiting digital content owners.	14	LIM-LEG-IP1; LIM-LEG-IP2; LIM-LEG-IP3; LIM-LEG-IP4; LIM-LEG-IP5; LIM-LEG-IP6; LIM-LEG-IP7; LIM-LEG-IP8; LIM-LEG-IP9; LIM-LEG-IP10; LIM-LEG-IP11; LIM-LEG-IP12; LIM-LEG-IP13; LIM-LEG-IP14
4. Competition. Refusals to share data may in certain cases raise abuse of dominant position concerns, especially as regards to leading technology companies (such as Google, Amazon, Facebook, Apple and Microsoft).	3	LIM-LEG-COM1; LIM-LEG-COM2; LIM-LEG-COM3
5. Surveillance. Surveillance, by public or private actors, enabled by aggregated Big Data analysis, threatens individual privacy. The threat of surveillance is likely to lead to further restrictions on the use of Big Data technology.	5	LIM-ES-SUR-1; LIM-ES-SUR-2; LIM-ES-SUR-3; LIM-ES-SUR-4; LIM-ES-SUR-5
6. Privacy and data protection. Compliance with the General Data Protection Regulation (GDPR) raises technical challenges regarding the principles of data	19	LIM-EP-RAD-1; LIM-LEG-DP1; LIM-LEG-DP2; LIM-LEG-DP3; LIM-LEG-DP4; LIM-LEG-DP5;

minimization, purpose limitation and transparency, relying on a data subject's consent, and responding to requests from data subjects.		LIM-LEG-DP6; LIM-LEG-DP7; LIM-LEG-DP8; LIM-LEG-DP9; LIM-LEG-DP10; LIM-LEG-DP11; LIM-LEG-DP12; LIM-LEG-DP13; LIM-LEG-DP14; LIM-ES-PRI-1; LIM-LEG-FF1; LIM-LEG-DIG6; LIM-LEG-DIG5
7. Anonymisation and pseudonymisation. While necessary to avoid the application of data protection legislation, anonymisation and pseudonymisation techniques limit the effectiveness of some Big Data applications. Nevertheless, the threat of reverting the processed data to personal data, remains.	7	LIM-LEG-AP1; LIM-LEG-AP2; LIM-LEG-AP3; LIM-LEG-AP4; LIM-LEG-AP5; LIM-LEG-AP6; LIM-LEG-AP7
8. Mixed datasets. Big Datasets commonly used in the transport sector are often composed of both non-personal data and personal data that are inextricably linked. The comprehensive GDPR will apply fully and without limitation to all data contained in such datasets, thus severely reducing the value of it.	4	LIM-ES-PRI-2; LIM-ES-PRI-3; LIM-LEG-FF2; LIM-LEG-FF1
9. Data quality. There is a high risk that the quality of transport data is low and unreliable. Extracting value in light of poor data quality increases the costs of processing, risks of error and failure, and hampers the growth of the data industry.	4	LIM-TM-DR-2; LIM-EP-IR-2; LIM-ES-FW-1; LIM-ES-TR-1
10. Data analysis. The process of data analysis transforms and models data to discover useful information needed in decision-making. This requires specialized techniques, tools and expertise. An organization must be prepared to invest for the long term in research to successfully and reliably create value from analysis.	3	LIM-TM-DC-3; LIM-TM-NT-4; LIM-EP-OLC-2
11. Social discrimination. Big Data analytics are susceptible to technical and systematic biases which can lead to discriminatory conclusions. These biases may be caused by data heterogeneity, the size of the data sets, data quality, noise accumulation, spurious correlation, incidental endogeneity, and the complexity of	6	LIM-TM-NT-5; LIM-EP-IR-1; LIM-EP-UPR-1; LIM-ES-SD-3; LIM-ES-SD-2; LIM-ES-SD-1

algorithms. These errors are often difficult to detect and correct		
12. Data sharing obligations. Data sharing obligations are increasingly adopted in the context of Intelligent Transport Systems. The adoption of data sharing obligations through legislation is however not necessarily the adequate way to increase data sharing, which in certain cases could be increased through alternative means such as public tenders or other incentivising schemes.	1	LIM-LEG-DS1
13. Data ownership. Despite the absence of specific ownership rights in data, multiple actors involved in the Big Data value chain could try to claim ownership of (or parts of) a dataset, which would hinder the production of, access to, linking and re-use of Big Data.	6	LIM-LEG-DO1; LIM-LEG-DO2; LIM-ES-DO-1; LIM-ES-DO-2; LIM-ES-DO-3; LIM-ES-DO-4
14. Data sharing agreements. Contractual agreements for data sharing cannot be enforced vis-à-vis third parties. This entails that no recourse is available against third parties that obtain unjustified access to or misuse the data.	5	LIM-LEG-DSA1; LIM-LEG-DSA2; LIM-LEG-DSA3; LIM-LEG-DSA4; LIM-LEG-DSA5
15. Liability. The current EU legal framework on liability entails legal uncertainty and accordingly stifles the uptake of Big Data in the transport sector.	3	LIM-POL-PU2; LIM-LEG-LI1; LIM-LEG-LI2
16. Data infrastructure. The infrastructure needed to store and use Big Data technologies is complex and costly. Organizations find it challenging and costly to select, plan, and use the right infrastructure and tools needed for the entire data value chain (e.g. collecting, storing, maintaining, analysing, etc.).	5	LIM-TM-DC-1; LIM-TM-LS-1; LIM-TM-NT-2; LIM-TM-NT-7; LIM-EP-IU-2
17. Data security. Protecting data from unauthorized use is becoming more complex and expensive. Data security, which involves data encryption, tokenization, and key management practices, is a highly specialized function that few organizations are equipped to handle.	11	LIM-LEG-CBO1; LIM-LEG-CBO2; LIM-LEG-CBO3; LIM-LEG-CBO4; LIM-LEG-CBO5; LIM-LEG-CBO6; LIM-LEG-CBO7; LIM-LEG-CBO8; LIM-LEG-CBO9; LIM-LEG-CBO10; LIM-EP-OLC-1

18. Data professionals. There is a shortage of data scientists, engineers, analysts, and developers in the local job market	2	LIM-TM-NT-1; LIM-EP-OLC-4
19. Data standardization. There are a variety of transport data formats, standards and technologies used within the transport sector, even by organizations providing the same services or by Member States across the EU. The lack of standardisation reduces collaborative research, large-scale analytics and the sharing of sophisticated tools and methodologies.	3	LIM-TM-DR-3; LIM-POL-PU1; LIM-POL-PR-2
20. Open data licenses. Although standard open data licences are encouraged by the PSI Directive (Directive 2003/98/EC on the re-use of public sector information), licences are still widely diverging across the EU. Any company that wishes to reuse PSI from different Member States with the aim of developing a product will have the technical and legal challenge of dealing with these licences.	5	LIM-LEG-PSI1; LIM-LEG-PSI2; LIM-LEG-PSI3; LIM-LEG-PSI4; LIM-LEG-PSI5
21. Change management. Actors in the transport sector are slow and unwilling to adapt their organization to incorporate the use of Big Data technology. Some issues are the unwillingness to experiment, to restructure the organization and business practices, and to realign and update the different internal data policies.	3	LIM-EP-OLC-3; LIM-EP-OLC-5; LIM-EP-IU-1
22. EU Big Data transport industry. The development of Big Data technologies in the EU itself lags behind other global leaders. The EU transport industry cannot benefit from synergies and co-development with the local data industry.	1	LIM-EP-IU-3

Appendix B: Questionnaire

Introduction

Big Data is a valuable resource that could enhance the economic sustainability and competitiveness of the European transport sector. However, the uptake of Big Data technology in the transport sector is influenced by many factors.

The following survey asks your opinion on 22 major Big Data-related issues. Your input is very valuable. It will be used to develop a policy roadmap for the European Commission, aimed at unlocking the full potential of Big Data in the transport sector.

We are asking you – as an expert - to evaluate each of the 22 issues on the severity of its impact. To do so, please consider each of the presented issues and select the most appropriate answer to the following question:

How limiting is this issue to the further application of Big Data in the transport sector?

It would be most helpful, if you also provide additional comments, justification, or disagreements in the space provided.

The survey takes about 15 to 25 minutes to complete. Please note that your responses are completely anonymous, and we require no details from you that would enable you to be identified amongst the responses we receive to the survey. By completing the survey, you are providing your consent for your responses to be used as part of this research. If you choose to provide us with your email address at the end, we will store this separately and securely.

The survey is part of the Leveraging Big Data in Transport Operations (LeMO) project, funded by the European Commission under the Horizon 2020 funding programme. For more information, please take a look at the project website at https://lemo-h2020.eu/. Specifically, if you would like background information to the issues, before taking part in the survey, please find our report at https://lemo-h2020.eu/newsroom/2019/10/1/new-report-published-identification-and-characterisation-of-barriers-and-limitations.

---

1. What type of organization do you belong to?
   - Goods and services
   - Transport sector
   - ICT sector
   - Public sector
   - Knowledge institute
   - Stakeholder organizations

1b. Please describe briefly the type of activities your organization does.

____free text

2. Please describe your organization's role, experience or interest with Big Data. Please select all the answers that apply.
   - We provide Big Data technology
   - We supply Big Data
   - We use Big Data
   - We research and develop Big Data technology and techniques
   - We research the economic and societal impacts of Big Data
   - We are interested in the regulation of Big Data
   - We want to preserve our members' interests concerning Big Data
   - Other --- free text

3. Please indicate where your organization operates, with respect to the countries of the European Economic Area (EEA), Switzerland (CH), and the United Kingdom (UK). Please select all the answers that apply.
   - Within one of the countries mentioned (domestic/national)
   - Across two or more of the countries mentioned (intra EEA+CZ+UK)
   - Outside any of the countries mentioned (extra EEA+CH+UK), please specify

4. Please describe the transport sector you are working with predominantly. Please select all the answers that apply.
   - None
   - Pedestrian or cycling
   - Road
   - Public transport
   - Rail
   - Waterborne transport
   - Aviation
   - Passenger transport
   - Freight transport
   - Other – free text

How limiting is this issue to the further application of Big Data in the transport sector?

1. Not limiting
2. Slightly limiting
3. Moderately limiting
4. Very limiting
5. Extremely limiting

6. No comment, insufficient information to judge

Collecting data

1. **Collecting data.** Gathering the right data to develop transport-related Big Data technologies remains virtually impossible. Transport data is often simply inaccessible, expensive to procure or collect, spread across multiple sources, or not available in digital form.
2. **Availability of open data.** Too few transport-related datasets are available as open data. While some are simply not published, although required to, some remain inaccessible due to bureaucracy, technical incompatibility, or unnecessarily restrictive licensing.
3. **Intellectual property rights.** The recent EU copyright reform introduced exceptions (covering both copyright and database right) to the exclusive rights of right holders. One of the exceptions will likely lead to text and data miners exploiting digital content owners.
4. **Competition.** Refusals to share data may in certain cases raise *abuse of dominant position* concerns, especially as regards to leading technology companies (such as Google, Amazon, Facebook, Apple and Microsoft).

Please provide comments, justification and disagreements (if any) for the answers you have selected above. (Optional) __freetext

Preserving privacy

5. **Surveillance.** Surveillance, by public or private actors, enabled by aggregated Big Data analysis, threatens individual privacy. The threat of surveillance is likely to lead to further restrictions on the use of Big Data technology.
6. **Privacy and data protection.** Compliance with the General Data Protection Regulation (GDPR) raises technical challenges regarding the principles of data minimization, purpose limitation and transparency, relying on a data subject's consent, and responding to requests from data subjects.
7. **Anonymisation and pseudonymisation.** While necessary to avoid the application of data protection legislation, anonymisation and pseudonymisation techniques limit the effectiveness of some Big Data applications. Nevertheless, the threat of reverting the processed data to personal data, remains.
8. **Mixed datasets.** Big Datasets commonly used in the transport sector are often composed of both non-personal data and personal data that are inextricably linked. The comprehensive GDPR will apply fully and without limitation to all data contained in such datasets, thus severely reducing the value of it.

Please provide comments, justification and disagreements (if any) for the answers you have selected above. (Optional) __freetext

Results of data

9. **Data quality.** There is a high risk that the quality of transport data is low and unreliable. Extracting value in light of poor data quality increases the costs of processing, risks of error and failure, and hampers the growth of the data industry.
10. **Data analysis.** The process of data analysis transforms and models data to discover useful information needed in decision-making. This requires specialized techniques, tools and expertise. An organization must be prepared to invest for the long term in research to successfully and reliably create value from analysis.
11. **Social discrimination.** Big Data analytics are susceptible to technical and systematic biases which can lead to discriminatory conclusions. These biases may be caused by data heterogeneity, the size of the data sets, data quality, noise accumulation, spurious correlation, incidental endogeneity, and the complexity of algorithms. These errors are often difficult to detect and correct.

Please provide comments, justification and disagreements (if any) for the answers you have selected above. (Optional) __freetext

Legal uncertainty surrounding data sharing

12. **Data sharing obligations.** Data sharing obligations are increasingly adopted in the context of Intelligent Transport Systems. The adoption of data sharing obligations through legislation is however not necessarily the adequate way to increase data sharing, which in certain cases could be increased through alternative means such as public tenders or other incentivising schemes.
13. **Data ownership.** Despite the absence of specific ownership rights in data, multiple actors involved in the Big Data value chain could try to claim ownership of (or parts of) a dataset, which would hinder the production of, access to, linking and re-use of Big Data.
14. **Data sharing agreements.** Contractual agreements for data sharing cannot be enforced vis-à-vis third parties. This entails that no recourse is available against third parties that obtain unjustified access to or misuse the data.
15. **Liability.** The current EU legal framework on liability entails legal uncertainty and accordingly stifles the uptake of Big Data in the transport sector.

Please provide comments, justification and disagreements (if any) for the answers you have selected above. (Optional) __freetext

Business model resources

16. **Data infrastructure.** The infrastructure needed to store and use Big Data technologies is complex and costly. Organizations find it challenging and costly to select, plan, and use the right infrastructure and tools needed for the entire data value chain (e.g. collecting, storing, maintaining, analysing, etc.).
17. **Data security.** Protecting data from unauthorized use is becoming more complex and expensive. Data security, which involves data encryption, tokenization, and key

management practices, is a highly specialized function that few organizations are equipped to handle.

18. **Data professionals.** There is a shortage of data scientists, engineers, analysts, and developers in the local job market.

Please provide comments, justification and disagreements (if any) for the answers you have selected above. (Optional) __freetext

Standards and licenses

19. **Data standardization.** There are a variety of transport data formats, standards and technologies used within the transport sector, even by organizations providing the same services or by Member States across the EU. The lack of standardisation reduces collaborative research, large-scale analytics and the sharing of sophisticated tools and methodologies.
20. **Open data licenses.** Although standard open data licences are encouraged by the PSI Directive (Directive 2003/98/EC on the re-use of public sector information), licences are still widely diverging across the EU. Any company that wishes to reuse PSI from different Member States with the aim of developing a product will have the technical and legal challenge of dealing with these licences.

Please provide comments, justification and disagreements (if any) for the answers you have selected above. (Optional) __freetext

Organization challenges

21. **Change management.** Actors in the transport sector are slow and unwilling to adapt their organization to incorporate the use of Big Data technology. Some issues are the unwillingness to experiment, to restructure the organization and business practices, and to realign and update the different internal data policies.
22. **EU Big Data transport industry.** The development of Big Data technologies in the EU itself lags behind other global leaders. The EU transport industry cannot benefit from synergies and co-development with the local data industry.

Please provide comments, justification and disagreements (if any) for the answers you have selected above. (Optional) __freetext

---

Though we have tried to make the survey as complete as possible, we realize that there may be some perspectives and issues we have neglected. If you have any further comments, for instance, on other high impact Big Data issues, which are not covered in the survey, please describe them below. (Optional) __freetext