LeMO

Leveraging Big Data to Manage Transport Operations

Research and Policy Roadmap for Successful Big Data Implementation in the Transport Sector

WESTERN NORWAY RESEARCH INSTITUTE
VESTLANDSFORSKING

Panteia
Research to Progress

Bird & Bird

CORTE
Confederation of Organisations in Road Transport Enforcement

This roadmap has been produced as a deliverable (D.4.4) within the framework and scope of the EU funded Horizon2020 project Leveraging Big Data to Manage Transport Operations (LeMO). The project explores the implications of using big data to enhance economic sustainability and competitiveness of European transport sector. It studies and analyses big data in the European transport domain looking at various aspects, including technological, legal, ethical, social, environmental, economic, policy and political. LeMO has also conducted a series of case studies to develop the recommendations contained in this report.

October 2020

Distribution level	Public (P)
Authors	*Tharsis Teoh, Panteia B.V.* *Ivo Hindriks, Panteia B.V.* *Remy Russotto, CORTE* *Jayant Sangwan, CORTE* *Isis de Moortel, B&B* *Minsung Hong, WNRI* *Rajendra Akerkar, WNRI*
Reviewers	*Christina Nash, CORTE* *Marusa Benkic, CORTE* *Svein Ølnes, WNRI*
Keywords	*Big data, research roadmap, policy roadmap, policy actions, research actions*

Disclaimer
This report is part of the LeMO project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 770038.
The content of this report reflects only the authors' view. The European Commission and Innovation and Networks Executive Agency (INEA) are not responsible for any use that may be made of the information it contains.

Executive summary

This deliverable presents the final version of the LeMO Roadmap. The proposed roadmap includes a set of actions from the research and policy point of view that governments, businesses and industries in Europe should implement in order to effectively take advantage of big data solutions with the aim of providing added value for society in the transport sector.

It is the culmination of work completed under the previous tasks of the LeMO project and is an extraction of the key outcomes from the first version of the roadmap. During the preparation of the roadmap, a validation workshop was held to obtain feedback on the draft roadmap, to validate its findings and to further extend the results to a broader range of stakeholders.

This roadmap is divided in two sections: namely a research roadmap focused on identifying the research priorities at stake, and a policy roadmap providing a set of recommendations to build a policy framework which will ensure that the European transport sector can make the most out of the data transition.

In total, the roadmap contains eleven research actions and thirteen policy actions, which are spread over eleven themes. These actions are designed to improve the data economy in the transport sector and provide a source of inspiration for further research in this area. This roadmap ends with a suggestion for the implementation of the actions.

The aim of the roadmap is to guide European transport policy and research efforts to develop an EU model for a socially responsible big data economy for transport, to allow stakeholders to identify and meet big data challenges and proceed with a shared understanding of the societal impact and concrete problems worth investigating in future programmes.

Table of Contents

List of Figures

List of Tables

Glossary

ABBREVIATION	EXPRESSION
BYTE	Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities
EU	European Union
GDPR	General Data Protection Regulation
GPS	Global Positioning System
IoT	Internet of Things
ITS	Intelligent Transport Systems
JSON	JavaScript Object Notation
LeMO	Leveraging Big Data to Manage Transport Operations
NOESIS	Novel Decision Support tool for Evaluating Strategic Big Data investments in Transport and Intelligent Mobility Services
OECD	Organisation for Economic Co-operation and Development
PSI	Public Sector Information
SFIA	Skills Framework for the Information Age
SME	Small and Medium-sized Enterprises
WP	Work Package
XML	Extensible Markup Language
Ontology	An ontology can be defined as a formal specification for the purpose of delimiting and grouping instances/concepts, based on their common class, and thus formalising a full or a subset of a domain.
Knowledge graph	A knowledge graph embeds data in the context through coupling and semantic metadata, this way it delivers a sustaining framework for data integration, alliance, analytics and distribution.
Linked open data	Linked Data is a set of design principles for sharing machine-readable interlinked data on the Web. When combined with Open Data (data that can be freely used and distributed), it is called Linked Open Data.
Dimensionality reduction	It is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data.

Introduction

This document presents the LeMO roadmap to capture the political, economic, social, technological, legal and ethical benefits associated with the use of big data in the European transport field. It provides the policy and research actions necessary to strengthen opportunities and diminish limitations associated with big data in the transport sector and the recommendations to assist industry, policy makers and scientists to address challenges in order to improve innovation and competitiveness.

The aim of the roadmap is to guide European transport policy and research efforts to develop an EU model for a socially responsible big data economy for transport. The roadmap allows stakeholders to identify and meet big data challenges and proceed with a shared understanding of the societal impact and concrete problems worth investigating in future programmes. The LeMO roadmap is expected to guide European policy and research efforts to develop a socially responsible big data economy in the field of transport and logistics.

LeMO methodology

The work presented here is the culmination of a series of case studies, analyses, expert focus studies and workshops conducted within the LeMO project, managed within five work packages (WP) as illustrated in Figure 1. The LeMO project studied and analysed the application of big data in the transport sector with respect to five transport dimensions: mode, sector, technology, policy and evaluation. Initially a review and mapping exercise was conducted in WP1 and WP2 to delineate what constitutes "big data" in the transport sector. The LeMO project covers four Vs of big data with a focus on the transport sector[1].

Volume – This is the most popular and obvious characteristic of big data, as the first word of "big" is indicating this characteristic. The massive volume of data collected from millions of vehicles in Europe, social data, sensor data, etc.

Variety – The differences within the industry standards, sampling rates, and data types such as video, JSON (JavaScript Object Notation), XML (Extensible Markup Language), pictures, text and more.

Velocity – The high arrival rate of data, for instance, sensor data, weather data, GPS (Global Positioning System) generated data, social media messages, and data generated from vehicle onboard devices.

Veracity – The potential for missing or erroneous data due to environmental conditions, unreliable data sources, equipment failures or malicious intent.

WP1 involved the generation of a shared **understanding of the current big data landscape**, including relevant technologies, applications, initiatives, policies and infrastructures, whereas WP2 explored the various **political, economic, social, technological, legal and ethical issues associated with transport big data.**

These findings fed into the next phase of the project, which involved the exploration of present and future societal impacts relating to the use of big data. In WP3 a total of **seven case studies were conducted on big data practices in transport and logistics** to gain an understanding of the political, economic, social, technological, legal and ethical externalities involved in them.

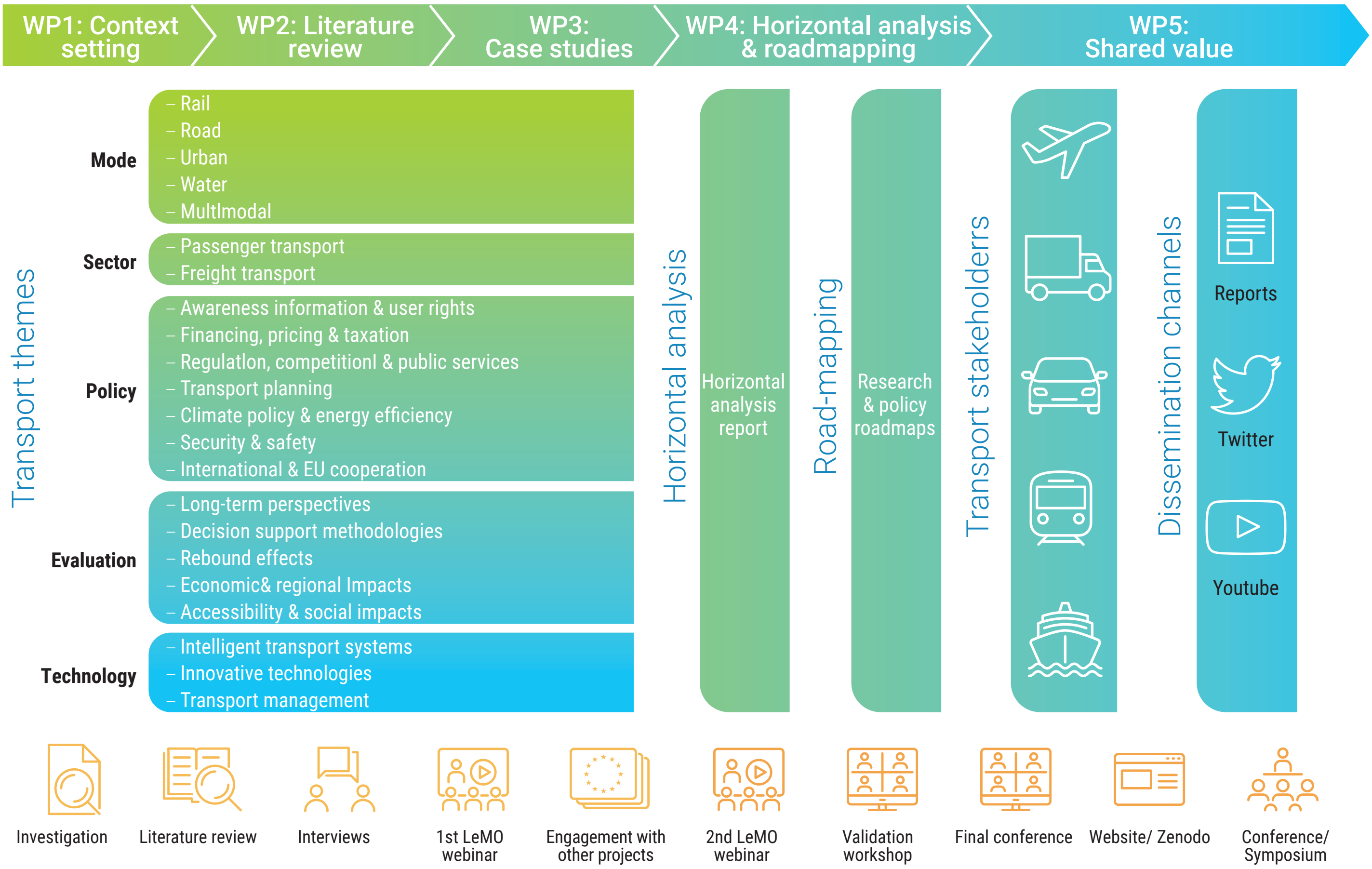
The outcome of these three WPs then fed into the final phase of the LeMO project, which involved establishing the direction to support European transport stakeholders in capturing a larger share of the big data market by appropriately addressing opportunities, challenges and limitations associated with transport big data. To achieve this, a **horizontal analysis of the societal facets encountered in the case studies** was conducted in WP4 to identify how they are connected to big data practices and to each other; with the aim to then evaluate and recommend how to address them.

Finally, the **key recommendations were extracted and further elaborated in a roadmap** which is presented in this document. The roadmap presents research and policy recommendations for identifying and addressing the societal impacts of big data, particularly amplifying the opportunities and diminishing the limitations associated with transport big data. During the preparation of the roadmap, two surveys were carried out among data experts to validate the findings from previous WPs and to further extend the results to a broader range of stakeholders. These outcomes have been further **validated and enriched by the LeMO Advisory and Reference Group,** which consists of experts from the data and transport domain.

During the course of the LeMO project, **engagement and dialogue with stakeholders** were key success factors for the realisation of a research and policy roadmap for big data in transport and to achieve LeMO's success. In WP5,

1 https://lemo-h2020.eu/s/20180711_D11_Understanding-and-mapping-big-data-in-transport-sector_LeMO.pdf

Figure 1: Overall methodology in the LeMO project



the knowledge generated within the project is implemented through interaction with stakeholders to create shared value.

Introduction to the roadmap

The roadmap consists of **two sections**: the **policy roadmap** and the **research roadmap.** The policy roadmap presents a set of policy actions necessary to develop a European transport big data ecosystem. The research roadmap detected research priorities to promote innovation. The policy roadmap guarantees the crucial conditions for innovation and the research roadmap pinpoints areas in which to capitalise. **Both the research and policy roadmaps present different perspectives which, when combined, outline a clear set of actions for European policy and research from the political, economic, social, technological, legal and ethical perspectives**.

An overview of the policy and research actions identified in LeMO and which are part of the roadmap are shown in figure 2. In total, eleven research actions are identified and thirteen policy actions, spread over eleven clusters. These actions are intended to serve as a guide to improve the data economy in the transport sector and as a source of inspiration for further research in this area.

This document is organised as follows. The research roadmap is presented first followed by the policy roadmap. Both roadmaps begin with an overview of challenges, followed by the contextualisation of the roadmap within LeMO, the larger policy space and research action in Europe. Each roadmap then outlines the key research and policy actions for Europe.

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;
- Research organisations;
- Industry;
- Authorities (regional and national level) that develop and enforce policies and legislation.

Figure 2: Overview of the 11 research actions and 13 policy actions

Research Roadmap

The research roadmap presents an action plan to carry out research related activities which support the use of big data technology and the growth of the data economy in the transport sector. The research roadmap lists challenges and actions to overcome these challenges. The roadmap targets a range of actors including (but not limited to) researchers, academics, data and transport practitioners, whether in the private or in the public sector. The challenges and actions listed hereunder can also serve as an indication to entities that fund research; providing them a basis to discern priorities and focus their attention on areas that can help foster the systematic application of big data technologies in the transport domain.

The activities under the LeMO project have been carried out in close collaboration with other Horizon 2020 research projects funded under the same topic, namely the Novel Decision Support tool for Evaluating Strategic Big Data investments in Transport and Intelligent Mobility Services (NOESIS)[2]. In view of this, the research roadmap both builds and expands the Technological roadmap proposed under NOESIS[3]. The challenges and actions proposed below should be seen as complimentary and additional to the four high-level technological actions proposed under the NOESIS project. The research roadmap has also taken into account the work done under other projects such as the BYTE project[4], Big Data Policy Canvas Project[5] and the strategic research agenda of Big Data Value Association (BDVA)[6] and EU[7].

Challenges

The LeMO research roadmap addresses the following key research-related challenges that have been identified through the work done before.

Semantic interoperability: Data (semantic) interoperability in storage, exchange and integration of transport data is vital for efficient, user-driven or automated, annotations and transformations.

Data quality assurance: Increasing quality of transport data is essential for improving trust in the results of data analysis. Using high-quality data is the first step towards extracting the right value from big data.

Data standardisation: Many stakeholders have pointed out that a lack of transport data standards is a significant obstacle. Data standardisation is crucial for data exchange and interoperability. However, addressing this problem is difficult due to the variety of data types.

Scalable and real-time data analytics: Scalable and real-time analysis is a significant challenge for timely extracting valuable information from big data that is generated continuously and in large quantities in the transport domain (e.g. constant stream of traffic data).

Event and pattern discovery: Identifying patterns from transport related big data is useful to gain both macro and micro level perspective and to detect trends. The patterns can also help recognise or forecast events such as an accident or machine defect.

Interactive and dynamic data visualisation: Interactive and dynamic visualisation tools can be helpful for easily understanding the outcomes of data analysis and for making appropriate decisions in shorter timeframes. Even simple visualisations can help attain clearer insights and facilitate deeper analysis.

Big data storage with edge computing: Big data storage (such as Cloud) is necessary for advanced analyses of large datasets. However, in sectors such as transport, where various types of data are generated rapidly and in huge quantities, it may not be feasible to store, manage and process all data in a stand-alone system. Edge computing may help such systems to analyse huge amounts of transport data by moving computation needs to sub-level components (edge and fog systems).

Infrastructure for real-time application: Existing infrastructures for Intelligent Transport Systems (ITSs) are not sufficiently designed for processing and analysing gathered data in (near-) real-time. Besides, they do not meet the emerging requirements for critical applications, such as low latency.

2 https://cordis.europa.eu/project/id/769980
3 https://drive.noesis-project.eu/index.php/s/cNup8ZG1pUL7BCG
4 https://cordis.europa.eu/project/id/619551
5 https://cordis.europa.eu/project/id/769623
6 https://www.bdva.eu/
7 https://ec.europa.eu/jrc/en/publication/strategic-transport-research-and-innovation-agenda-stria-roadmap-factsheets

Skilled workers/experts (with domain knowledge): Currently there is an immense shortage of professionals with big data analytics skills. In the transport field, people with big data analytics know-how are further required to have the knowledge of transport domain and its specificities, which adds another layer of complexity.

Robust data protection: Since some big data techniques can extract personal information; transport data, in the context of privacy, requires more robust data protection in order to bring itself into compliance with the General Data Protection Regulation (GDPR). Nonetheless, such an increase in protection has a trade-off relationship with the availability of data.

Cybersecurity: Increasing digitalisation, connectivity and data in the transport domain is also increasing its exposure to cyber security threats. In addition to conventional security threats, there is also a need to consider new types of security threats, which may arise on account of misuse, alteration or misrepresentation of data by a threat actor.

Research actions

For each challenge described above, corresponding actions have been proposed, which have been identified through research and discussion with experts. The first part of the description of each cluster discusses the actions, while the second part provides an overview of expected results and impacts of each action.

Data management and interoperability

Research on data management and interoperability enable the transport sector to manage and leverage big transport data effectively and efficiently while assuring data quality. There are research recommendations on three significant issues under this cluster.

Research action 1 Semantic interoperability

Research on topics such as "semantic search and matching", "ontology", and "linked data", should be conducted to improve interoperability between transportation data storages and systems. Transport domain experts and stakeholders should be actively involved in the research, to comprehensively express the data structure of their domain knowledge (on transport data properties and their relationships) as an ontology, and the data should be transformed into a knowledge graph. An automatic and interactive tool, converting the domain knowledge of transport experts into an ontology, could also be developed. To accomplish this, the following areas need to be addressed:

- Research on ontology and knowledge graph for the transport sector;
- Development of interactive tools to convert transport domain knowledge.

Research action 2 Data quality assurance

Algorithms and data quality management frameworks must be developed to address a wide range of data quality related problems by filtering outliers, estimating data uncertainty, grouping data into clusters and comparing relevant data. These would help improve, measure, and validate data quality. Close collaboration with transport domain experts would be crucial here, to ensure the expected performance of these algorithms for the transport sector. To accomplish this, the following areas need to be addressed:

- Design and development of data quality management frameworks;
- Research and development of better techniques for the frameworks;
- Encourage the involvement of transport stakeholders in the development process of the techniques and frameworks.

Research action 3 Data standardisation

To disambiguate data and extract representative factors; "dimensionality reduction" and "normalisation" must be employed. This can be achieved by leveraging "ontology" and "linked open data" to semantically link and classify semi-(un)structured data clusters. A collaborative platform or a community needs to be established to gather knowledge from transport domain experts and practitioners to set standards based on utility and harmonise rules and practices to achieve greater interoperability. This action is closely related to policy action 10.

To carry out this action, the following areas need to be addressed:

- Development of a collaborative platform deriving harmonisation of rules and practices;
- Establish and activate a collaborative community to collect domain knowledge.

Table 1: Results, barriers and impact of data management and interoperability research actions.

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Semantic interoperability	Various ontologies for the structures of big transport data. Various knowledge graphs of transport big data. Interactive tools for converting transport domain knowledge into the ontologies and knowledge graphs.	It may be difficult for transport stakeholders to convert their domain knowledge (Carenini et al., 2018). The need for validation between various ontologies in the same categories (Katsumi and Fox, 2019). The efficacy of tools needs to be demonstrated before the stakeholders adopt these tools (Ganzha et al., 2017).	Providing a better understanding of big transport data. Enabling stakeholders to contribute to semantic data standardisation through the tools directly. Improved interoperability in transport domain across EU, benefitting both service providers and recipients. Increased collaboration between data experts and transport domain experts.
Data quality assurance	Better algorithms for data quality problems specific to the transport industry. Consolidated frameworks including the algorithms for data quality management.	Difficulty to validate algorithms and frameworks as well as data quality problems (Megler et al., 2018). The collaboration with transport domain experts is essential for the validation (Luo et al., 2019).	Improving transport data quality and maximising the data value (Luo et al., 2019). Obtaining and guaranteeing the trust in transport data.
Data standardisation	A collaborative community for collecting transport domain knowledge. A collaborative platform for data analysis to derive harmonisation of rules and practices. Creation of more harmonised rules and practices across Europe to support data standardisation.	The analysis of different norms and practices across Europe. Contributions from transport stakeholders to validate the data analysis platforms (Chinrungrueng et al., 2018).	Data standardisation, which is also reinforced through the use of semantic techniques (i.e., ontology and knowledge graphs). Automatic data standardisation with minimal use of human resources (Seedah et al., 2016). Increase in data interoperability, ultimately leading to homogeneity in the transport domain across EU and fostering data sovereignty.

Data processing and analytics

Data processing and analytics play an essential role in extracting value from big transport data and in helping transport practitioners make correct and timely decisions. There are research recommendations regarding three significant issues under this cluster.

Research action 4 Scalable and real-time data analytics

Efficient and effective algorithms must be developed that can split/integrate data and the results of its analysis according to the architecture of edge computing. Such an algorithm should focus on avoiding fragmented analysis results and transferring them along with appropriate data to the upper systems (i.e., Cloud/Fog system layers). To accomplish this, the following areas need to be addressed:

- Development of scalable and real-time data analysis algorithms;
- Research on extreme analysis for analysing the transport network.

Research action 5 Event and pattern discovery

The development and deployment of intuitive tools, which offer data mining, statistics, and machine learning techniques, will help transport practitioners in identifying relevant patterns and in achieving specific transport objectives. To accomplish this, the following areas need to be addressed:

- Design and development of intuitive tools for identifying events and patterns;
- Pre-requirement analysis in the transport industry for these tools.

Research action 6 Interactive and dynamic data visualisation

Dynamic and interactive visualisation tools should be developed, specific for the needs of transport practitioners that can even be employed by non-technical users. Such tools should specifically focus on temporal and spatial data-driven visualisation, which can benefit the user interfaces of parallel exploration in the transport domain. To this end, inspiration can be drawn from tools such as Tableau[8]. To accomplish this, the following areas need to be addressed:

- Requirement analysis in the transport sector for visualisation tools;
- Development of visualisation algorithms and tools;
- Deployment of guidance and education for using the tools.

Table 2: Results, barriers and impact of data processing and analytics related actions research

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Scalable and real-time data analytics	Algorithms for scalable and real-time data analysis on the architecture of edge computing. Extreme analytics algorithms for autonomous transport networks (Oussous et al., 2018).	Infrastructure shortage for edge computing in the transport area (Asensio et al., 2020). Immature level for applying edge computing in the transport industry (Ferdowsi et al., 2019).	Analysis of more comprehensive information (e.g., trends) via scalable data analysis. Enabling transport services based on big transport data available in real-time. Real-time data analysis leading to quicker decision-making.
Event and pattern discovery	Intuitive tools for identifying patterns to achieve specific transport sector objectives.	Pre-requirement analysis of the transport industry is essential (Torre-Bastida et al., 2018). The need for collaboration with transport stakeholders (Torre-Bastida et al., 2018).	Faster event and pattern detection in the transport sector (Qin et al., 2018). Faster delivery of services and reduced response time. Improved decision and policy making, supported by accurate data analysis.
Interactive and dynamic data visualisation	Dynamic and interactive visualisation tools. Visualisation with multi-dimensional data.	Need of a preliminary analysis of practical requirements of the transport domain (Schoedon et al., 2019).	Increased uptake of sophisticated visualisation tools in the transport sector. Transport data analytics from multiple views and perspectives. More accurate, data-driven decision and policy making for the transport sector.

Infrastructure and skills

Infrastructure and skilled workers are crucial for adapting and applying big data technologies in the transport sector. A lack of these capabilities can severely impact EU's competitive advantage and the benefits that transport sector can reap from big data technologies. There are research recommendations on three significant issues under this cluster.

Research action 7 Big data storage with edge computing

Recent hierarchical approaches (such as Cloud/Fog/Edge computing) are to be applied to analyse data and send only the required data to the Cloud, instead of transferring all the data to it. This would reduce the burden on the Cloud system, making it possible to process and manage data in real-time. To accomplish this, the following areas need to be addressed:

- Benchmarking best practices of big data storage from other domains (e.g., IoT);
- Development of big data storage based on edge computing for the transport sector.

8 https://www.tableau.com/

Research action 8 Infrastructure for real-time application

Design systems, which consider multifaceted issues of big data, such as real-time processing and analytics, data veracity and variety. Cloud/Fog/Edge infrastructure and computing must be considered as alternatives, while being mindful of concerns related to cyber security and data protection. To accomplish this, the following areas need to be addressed:

- Requirement analysis of the transport infrastructure based on edge computing;
- Construction of sub-components of the infrastructure for fog and edge computing.

Research action 9 Skilled workers with domain knowledge

New online courses for re-skilling and up-skilling professionals and building an interdisciplinary workforce needs to be introduced. Such courses, with easy accessibility, will stimulate people to learn about the use of big data and to adopt new data-related skills. Developing a tool for collaborative education is essential for encouraging knowledge-sharing between domain practitioners and technicians, as well as students. To this end, the following areas need to be addressed:

- Developing online collaborative courses for big data technology for the transport sector;
- Incentivising stakeholders to participate in such courses.

Table 3: Results, barriers and impact of infrastructure and skills related research actions

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Big data storage with edge computing	Big data storage for edge computing. Hierarchical approaches based on hybrid computing (cloud and edge) (Jaisimha et al., 2020).	Need to handle dynamic and uncertain vehicular environment (Li et al., 2019). Increasing security and privacy threats on the distributed systems (Li et al., 2019).	Reduced load on Cloud system (Jaisimha et al., 2020). Real-time data management and processing that helps quicker decision-making.
Infrastructure for real-time application	The infrastructure for sub-level components (edge and fog nodes) in the transport sector. Real-time system considering security and data protection (Munir et al., 2019).	A lack of appropriate methods to measure security of transportation infrastructure. The difficulty of proper validation of measurements (Sun et al., 2020).	Encouragement and quantification of security and data protection on transportation infrastructure. Fostering edge computing for real-time applications.
Skilled workers /experts (with domain knowledge)	Online training courses leading to skilled workers and experts. A collaborative education tool for sharing transport domain knowledge.	Satisfying the needs of all types of learners coming from different backgrounds to attend the programmes and retaining them to complete the courses (Chalkiadakis et al., 2019). Difficulties in addressing various issues related to quality of the programme, collaboration among students, cultural differences, technological fit, and so on (Aparicio et al., 2019).	Skilled workers/ experts in the transport sector. Increased employment. Greater interaction between experts from different domains. Development of more data driven applications, ultimately improving transportation across the EU.

Data protection and security

The importance of data protection and security cannot be understated, especially with increasing interconnectedness and reliance on data technologies in the transport sector. Protection needs to be ensured against various conventional and emerging threats through research and awareness. There are research recommendations on two significant issues under this cluster.

Research action 10 Robust data protection

A method that quantifies privacy loss and data utility must be developed to deploy more robust and dynamic data protection techniques depending on the situation and type of data used (i.e. whether it is personal data or non-personal data). Several data protection techniques are available, each with their own pros and cons. For example, "anonymisation" has a lower cost than "encryption", but its reversibility may expose exploitable weaknesses. Also, there is a need to develop methods that can help evaluate the value that can be extracted from data as well as identify real risks in contrast to hypothetical risks. Additionally, data protection techniques under the various GDPR approaches, that foster the use of targeted data for concrete well-defined and limited purposes, should be expanded and promoted to ensure adequate data protection as well as data sovereignty, by awarding greater control to citizens over their data. Addressing the following areas can support this:

- Requirement analysis for data protection techniques pursuant to the GDPR;
- Development of data protection methods pursuant to the GDPR.

Research action 11 Cyber security

Research on topics such as "encryption", "threat intelligence", "differential privacy" and "assuring data reliability" in digital platforms and services needs to be reinforced. Furthermore, awareness of personnel working in the transport sector needs to be increased by developing courses and training programs to better understand threats, vulnerabilities, prevention and response protocols is necessary. From a technical perspective, blockchain technology can be leveraged in the transport domain to improve data transparency and trust. In summary, the following areas need to be addressed:

- Foster better understanding of conventional and emerging cyber security threats as well as corresponding response measures;
- Raising awareness among transport practitioners through awareness campaigns and training.

Table 4: Results, barriers and impact of data protection and security related research actions

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Robust data protection	Robust data protection methods that ensure adequate value extraction from data without compromising privacy and security. Data protection techniques using GDPR methods (Stefanouli and Economou, 2018).	A lack of harmonisation within the EU and cooperation between EU Member States (Stefanouli and Economou, 2018). Ensuring data protection thoughout the data life cycle (Li et al., 2018).	Development of dynamic and customised data protection techniques for the entire data life cycle (Li et al., 2018). Demarcation of hypothetical and real risks of using of big transport data (Westerlund, 2018). Maximum value extraction from data, while ensuring privacy. Prevention of data misuse and increase of data sovereignty.
Cybersecurity	Cyber security techniques considering "threat intelligence (Tucci, 2017)", "differential privacy (Yang et al., 2020)" as well as "assuring data reliability". Cyber security techniques for new and conventional threats. Awareness raising materials and training courses for transport practitioners.	Cyber security threats emerging from new mobility innovations (e.g., autonomous vehicles) (Axelrod, 2018).	Improved transparency and trust in processing big transport data (Sun and Song, 2017). Reduction in security threats due to human factors (CAIRE, 2017). Prevention of loss in terms of life, property, infrastructure, finances etc.

Policy Roadmap

Policy makers can carry out a set of actions to support the use of big data technology and the growth of the data economy in the transport sector. Our analysis has resulted in thirteen policy actions which are most effective in achieving this goal. Policy actions are defined broadly, referring to any action that can be taken by executive or law-making institutions. This could involve actions to enable and support a desired outcome towards citizens and organisations, but also towards internal functioning of governmental organisations. These actions must be understood as any guidelines, fiscal measures, regulations, marketing and communication strategies and service provision, i.e. any policy that enables or supports a desired outcome. Depending on the scope, the actions can be implemented at local, regional, national and European level — although most actions require implementation on a European level.

Challenges

The LeMO policy roadmap will address the following challenges. These have been refined through an extensive research and stakeholder engagement process. Full descriptions of the challenges are presented in D4.1.

Availability of data: The development of successful big data applications relies on the collection and acquisition of high quality data. Nevertheless, in many applications, collection of the right transport data remains a challenge. The data may prove simply inaccessible, expensive to procure or collect, is spread across multiple sources, or is not available in digital formats. In addition, the current ability of companies holding high value datasets to merge (or be acquisitioned) leads to unfair competition.

Data quality: While many good big data applications are built using the current state of data quality, it is insufficient for the application of more advanced techniques, especially in high impact, high risk and real-time settings. Poor data quality increases the cost of processing, risks of error and failure, and hampers the growth of the data industry.

Conflict and uncertainty within the legal framework: Practitioners require a clear legal framework, however there are some crucial points that need to be improved. In the absence of clear and specific data ownership rights, there is a tendency for value chain actors to claim ownership of data, thus halting development of the application. Further, the EU framework on liability and data sharing enforcement vis-à-vis third parties is inadequate and is an obstacle for the collaboration needed for applications of big data in the transport sector to flourish. However, while addressing the legal framework, policy makers must consider that more, and often more complicated, regulation may favour large companies with departments of legal competence. As such, control of data will be even more unbalanced than it is now.

Privacy challenges. Privacy is an important value that should underlie all big data applications. Many challenges to privacy posed by big data and the new data economy have been solved to an extent by recent EU regulations. While beneficial to the preservation of privacy, these efforts may hinder the widespread application of big data technology in the transport sector by posing technical challenges in data minimisation, purpose limitation, transparency, reliance on consent, and responding to requests from data subjects, as well as in dealing with mixed datasets.

Common standards for data technology: The lack of common data technology standards across the EU increases difficulties to promote big data technology application in the transport sector. The lack of standardisation reduces collaborative research, large-scale analytics and the sharing of sophisticated tools and methodologies.

Lack of talent and unprepared organisations: The two main challenges that hold back organisations specifically are: the lack of data scientists and professionals competent to develop big data applications, as well as the unwillingness of organisations to upgrade their operations to use the technology.

Policy actions

In total, thirteen policy actions are identified organised under seven clusters. The first part of each cluster discusses the actions, while the second part provides an overview of expected outcomes and impact of each action.

Ensuring sufficient talent

Transport organisations that are integrating big data analytics into their processes need to deal with the current lack of expertise and competence within their own companies and in the sector. Organisations have multiple ways to increase human resource capacity if it currently lacks the competence.

Policy action 1 New employment models

Gig economy platforms can help to fill the gap in personnel in the very short term, but new methods of working together securely, effectively and creatively must be developed and researched. Technological means such as open collaborative platforms are required. These collaborative platforms must be easy to use, combine communication and information sharing services and enable remote working on the same physical systems (i.e. data storage servers, analytics etc.) so that employees can carry out their tasks in a convenient and timely manner. However, teleworking options have their disadvantages and risks and should be implemented cautiously. Labour rates might be slightly higher than an individual employee, quality assurance procedures may not be currently suited to remote workers and freelancers, and continuity and loyalty for long term projects may be lacking. Data scientists need domain knowledge and close cooperation with operations and the business functions that are being transformed. In many cases, especially in the exploration of potential, regular communication and interaction within the team is necessary. Hence, good teleconferencing technology and other co-creation platforms would also be a good research opportunity. To accomplish this, the following areas need to be addressed:

- Good practices for teleworking;
- Open collaboration platforms for dynamically establishing collaborations;
- Marketplace and matchmaking tools;
- Appropriateness of regulatory framework for the gig economy.

Policy action 2 Training and education

Developing home-grown talent is important for ensuring big data innovations in the transport sector in the long term. Therefore, training and education programs need to be developed which are ready for the future skills demand. An understanding of the key skills needed in the future should guide the training programs designed now. The SFIA Foundation has developed a framework, the Skills Framework for the Information Age, for the types of skills that are needed to support the digital age[9]. Some of the relevant key themes are cyber security, digital transformation, big data and informatics. Universities, industries and the government should work together to develop integrated skills focused programs. Programmes must be developed for new students but also for re-training current workers in the transport sector. Including learning tools such as gamification can increase the attractiveness and effectiveness of these programs. Integrating internships and student projects with the industry into the programs will facilitate the transition to the general business world. The following measures can contribute to the implementation of this action:

- Develop a skills-framework for the big data skills needed in the European transport sector;
- Develop training and education programmes or learning tools specifically aimed at supporting the usage of big data in the transport sector;
- National and local governments can work closely with universities to develop internships and student projects for education purposes, to attract new talent and to prepare students for the business world;
- Provide scholarships to relevant education programmes.

Table 5: Results, barriers and impact of policy actions related to ensuring sufficient talent.

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
New employment models	Development of networks of new partners for collaboration. New platforms for open collaboration, providing Service/business capabilities repository. matchmaking, assessment of trustworthiness of a stakeholder. New platforms for teleworking allowing employees to carry out their tasks in a convenient and timely manner.	It is difficult for organisations to change existing ways of working. Advantages of collaborative platforms needs to be clearly demonstrated before organisations adapt this way of working.	Better matching of supply and demand of workers. Organisations have better access to human resources.

9 SFIA Foundation, https://www.sfia-online.org/en/framework/sfia-7/about-sfia-7

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Training and education	New and improved education programs that meets future skill demand. Students will gain more practical experience by collaboration with national and local governments.	Developments in big data technologies are going fast, so it is unforeseeable which skills will be needed in the future.	Better match between study programmes and the necessary knowledge in the business world. Easier transition of students into the business world.

Data management guidance

Data collection and processing activities typically involve the interaction of a multitude of actors. Determining the 'controllers' and 'processors' of data and their associated rights is often a complex exercise, as is putting in place appropriate contracts that cover all possible situations with the necessary and satisfactory legal certainty. One way to tackle this would be to provide guidance for the sector at a European level on how to handle these legal issues in the transport sector. In this respect, two actions have been included in this roadmap which are seen by organisations as the biggest obstacle to the application of big data, namely in the field of privacy and data protection, and in the field of data sharing.

Policy action 3 Data management guidance on privacy data protection

The legal analysis of privacy and security aspects in the context of big data is rather complex. Under the GDPR, the concepts of 'personal data' and 'processing' are defined and interpreted very broadly, to such an extent that the numerous obligations under the GDPR when performing big data analysis will apply in a broad range of circumstances, including in the context of the transport sector. Certain principles and requirements related to privacy and data protection are nevertheless difficult to fit with some of the main characteristics of big data analytics. In this respect, finding a balance between the various rights at stake is of paramount importance. It follows that the competent authorities at EU and national levels should develop guidance in this field. Guidance is needed in determining the applicability of the GDPR to various big data analytics models and to the precise roles of those involved in complex data processing activities. Furthermore, the development of template agreements, compliant with the requirements of the GDPR, would be a welcome evolution. This would increase legal certainty for those involved in the data value chain, and ultimately benefit data subjects. In doing so, it is essential to keep in mind that the right to the protection of personal data is not an absolute right, but must be considered in relation to its function in society and be balanced against other fundamental rights in accordance with the principle of proportionality. To accomplish this, the following areas need to be addressed:

- Develop additional guidance and template agreements on the usage of personal data that help to clarify the relationships between different actors in the data value chain in certain situations;
- Tailor guidance and templates to the intricacies of big data analytics in the transport sector.

Policy action 4 Guidance on data ownership and sharing

With many actors involved in the data value chain taking up various roles, it is often difficult to establish who has which specific rights in the data and how an actor is able to assert those rights. Different actors may claim ownership of (or parts of) a dataset, which could hinder the production of, access to, linking and re-use of big data, and consequently hamper innovations in the transport sector. The legal analysis undertaken in LeMO has demonstrated that owing to the absence of legislation on data ownership on a European level, the current legal framework is highly uncertain. At the time of writing, the only solution is to lay down the various rights of data in contractual agreements, often data sharing agreements. But such agreements usually do not cover all aspects related to rights attached to data, the way in which such rights can be exercised and by whom, specifically in relation to third parties. As such, a more legally secure solution is necessary. Policy makers can carry out this action in the following areas:

- Provide data sharing agreement templates tailored to the specific data needs and complexities of the transportation sector;
- Introduce EU legislation to create a non-exclusive, flexible and extensible ownership rights in data.

Table 6: Results, barriers and impact of data management guidance related policy actions.

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Data management guidance on privacy and data protection	Increased usage of personal data by organisations due to increased legal clarity. Non-action by organisations caused by uncertainty of what constitutes personal data is avoided.	Some interpretations provided by certain authorities are conservative, too restrictive, or simplistic to such extent that it would prohibit certain technologies of processing activities.	Data subjects are less suspicious of organisations using their personal data.
Data management guidance on data ownership and sharing	Increased data sharing due to more legal certainty about the rights in data and how to assert these rights.	Impracticality to create EU-wide legislation on rights in data, e.g. a data ownership right, due to the complexity of the roles and rights of stakeholders in the data value chain.	More willingness of organisations to collaborate and share transport data (sets) due to increased legal certainty.

Data sharing platforms

Policy makers may support the data economy and the uptake of big data technology by supporting data sharing platforms. Here, at least two different initiatives should be distinguished. First, a platform for selling and sharing data between any type of actor, whether individual, commercial, or public. The actors are not mandated or obliged to sell or share this data. The second initiative builds upon the open data initiatives based on government regulation.

Policy action 5 Data marketplace

A data marketplace facilitates data transactions between data suppliers and buyers. As with other marketplaces providing digital content, such as ebooks, music and movies streaming, the benefits are primarily the reduction of effort and inconvenience of finding the right data source, ensuring quality big data transmission, and facilitating revenue collection for data suppliers. Actors that only want to sell data "on the side" will find the marketplace supports their commercialisation efforts.

Specialised transport data marketplaces would cater to niche groups providing different levels of sophistication based on their big data specification needs. For instance, high volume and velocity data will require different needs from data sets with high variety and volume.

To support the emergence of transport data marketplaces, public authorities and policy makers can strive to:

- Resolve legal uncertainty regarding sharing of transport data, especially with regards to privacy concerns, data ownership, and GDPR compliance;
- Promote open source data marketplace software developed for open data portals;
- Participate as public entities in the marketplace as suppliers of high quality datasets or purchasers;
- Develop EU-wide data sharing marketplaces for urban mobility data (e.g. covering vehicle data, infrastructure use information, charging service location and use) and freight transport data (e.g. truck parking use, types of goods, vehicle location/density).

Policy action 6 Big open data initiatives

The public sector, from local to national to EU-level, possess a wide variety data, which is potentially valuable to the transport sector. Regulation has been instrumental in increasing the amount of data provided to the public in the open data approach. These efforts will need to increase with a focus on providing high quality data and supporting the technical specifications of big data.

Current open data portals focus primarily on static and highly aggregated data, which provide little value to big data applications. The transition that will be taken at the EU-level by providing "high-value datasets"[10] will be a strong step towards supporting innovative transport big data solutions.

To further develop this initiative, the following actions will need to be taken:

10 https://www.europeandataportal.eu/en/highlights/high-value-datasets

- Support the provision of big data using the open data approach, with guidance, research, and funding;
- Technical upgrade of the EU's open data portal into supporting the transmission of big data;
- Incentivise compliance with open data directives, especially with regards to the high value datasets. Additionally, provisions for liability and quality control in open data regulations should be added.

Table 7: Results, barriers and impact of policy actions related to data sharing platforms.

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Data marketplace	New data market places for various niches. Increased synergy between open data platform and data market places. Increased perception of value of transport data.	Resolve legal uncertainty and technical challenges regarding sharing of transport data. Marketing of data marketplaces. High cost of maintaining infrastructure.	Increased financial incentive to produce and supply high quality data. Increased merging of big data sets from multiple organisations leading to more innovative transport solutions.
Big open data initiatives	Increased awareness of big data needs and opportunities for open data. EU's open data portal ready for big data.	Overcome public sector inertia and lack of awareness with regards to big open data. High cost of providing high quality open data sets.	Lowered barrier to provide big data applications. Higher innovation expected. Support for individuals and start-ups to test and develop big data applications.

Technology development

The EU can support the transport sector by encouraging the development of technology for rights in data, such as for instance data ownership, and for data protection. The objective of these actions will be to increase the availability of usable data, while ensuring societal and commercial values, such as privacy, security, and intellectual property rights, will be preserved.

Policy action 7 Techniques for rights in data including data ownership

Some form of data ownership right could help to ensure that data suppliers are incentivised to "supply" data. Cryptographic technologies, such as blockchain, zero-knowledge proofs and public-key cryptography may provide users and businesses with functional control and ownership of their data. Blockchain and big data technology could be merged to provide decentralised but controlled access to high fidelity data, thus increasing (controlled) data sharing, while ensuring data integrity. Smart contracts could be used to implement complex ownership rights on the blockchain-based big data itself. Such technologies would eliminate the need for complex administrative and legal work to enable sharing and data ownership, of which the impact could be potentially huge.

The EU could support the development and application advanced technologies by the following:

- Providing guidelines on implementing cryptographic technology for certain high value database structures;
- Supporting open source cryptographic applications such as blockchain, especially for new mobility and transport applications;
- Integrating cryptographic technology in existing data portal structures.

Policy action 8 Techniques for data protection

To fully utilise high value transport datasets, while preserving privacy and commercial interests, the application of anonymisation, pseudonymisation, and data encryption techniques should be promoted. Suitable techniques that balance out the need for high quality and volume datasets are often required for analysis will need to be developed and promoted. A further description of techniques are provided in "**R10 - Robust data protection**".

Policy makers can support the adoption of the techniques by the following:

- Providing guidelines on the types of techniques suitable for transport big data applications;
- Incentivising the use of the techniques by subcontractors within the transport sector.

Table 8: Results, barriers and impact of policy actions related to technology development for data ownership and data protection.

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Techniques for rights in data including data ownership	Increased competence and awareness of cryptographic technology for transport big data needs. Growth of open source community for cryptographic transport applications. Existing data portals ready for cryptographic databases.	Lack of skilled developers in cryptographic technology. Inertia of organisations to implement new technology. The widespread decentralisation of data may bring about new challenges that needs to be researched.	Maintain data ownership and control while promoting big data applications. Higher confidence of data suppliers to provide data, if data ownership can be solved technically.
Techniques for data protection	Increased competence and awareness of data protection techniques used in transport big data applications. Increased adoption of data protection techniques by public sector subcontractors.	Lack of skilled developers of data protection techniques. No incentive to protect data, when access to databases are kept closed.	Higher level of data protection methods used increases trust in data sharing. Compliance with privacy regulations.

Data standardisation

Policy makers may support the data economy and the uptake of big data technology by supporting data standardisation. We distinguish two different approaches: harmonisation of data legislation on the one hand and supporting the development of data standards through standardisation bodies.

Policy action 9 Harmonisation of data legislation

The creation of a true common European data space will be difficult to achieve as long as the various data-related legislative frameworks across EU Member States remain fragmented. The lack of harmonisation at EU level could have a chilling effect on EU-wide big data projects, since it may require companies to comply with different types of requirement depending on the Member State in which a big data solution is deployed.

Harmonisation is therefore welcomed as this would create a more level playing field for companies wishing to deploy cross-border data-driven applications and technologies. The EU and other policy makers can support harmonisation of data legislation by:

- Conducting an assessment of which areas of legislation relating to data that currently have the biggest negative impact on innovation through lack of harmonisation.

Policy action 10 Data standardisation

Creating a common European data space among others requires addressing the challenge of having to work with a huge variety of data types, formats and quality levels. In the transport sector, as in many other sectors, this brings about the additional difficulty of interoperability. One approach to this challenge is to support the development of data standards through standardisation bodies. 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 could be tailored to the transport sector. This could facilitate the standardisation and increase the quality of transport sector data. The EU and other policy makers can facilitate data standardisation by:

- Supporting an effective and coherent data standardisation framework, which focuses on interoperability and high data quality;
- Encouraging the development of data standards through cooperation of authorities, standardisation bodies, service providers, vendors and industry players;

- Encouraging the development of certification mechanisms, seals and codes of conduct through cooperation of authorities, standardisation bodies, service providers, vendors and industry players and promoting their use across the European data space;
- Assessing whether competition law should be adapted to support data standardisation efforts;
- Setting requirements for data standards, including data quality and data interoperability, through public procurement.

Table 9: Results, barriers and impact of policy actions related to data standardisation

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Harmonisation of data legislation	Identification of key areas where harmonisation is lacking and required.	Chilling effect on EU-wide projects.	Lowered costs to provide big data applications. Increased innovation expected.
Data standardisation	Increased transparency through certifications and seals. Increased data interoperability through data standards. Increased data quality through data standards.	High interoperability costs due to huge varieties of data types and formats. Uncertainties regarding data quality.	Lowered costs to provide big data applications. Increased competition expected.

Data sharing in agreements

While private companies often generate huge amounts of data, they are not always prepared to voluntarily share this data outside the company. This is due to the large number of legal, commercial and technical challenges associated with private sector data sharing. We identified two distinct approaches to support data sharing by private companies by requiring them to share their data. First, data sharing obligations can be imposed as conditions in public tenders. Second, sector-specific or horizontal legislation can be adopted that imposes data sharing obligations.

Policy action 11 Data sharing obligations public tendering

An entirely different way of imposing data sharing obligations is by including them as conditions in public tenders. Public authorities can set requirements on the usage of data through public procurement, among others by setting requirements for data sharing. The creation of an open interface or platform suitable for big data applications by the private company chosen for the tender could for instance encourage start-ups and SMEs to develop innovative services by using the data that are thus made accessible to them. The EU and other policy makers can facilitate data sharing by:

- Setting obligations for data sharing through public procurement.

Policy action 12 Data sharing obligations legislation

Data sharing by private companies may also be supported by legally requiring them to share their data. Increasingly, technical specifications are required for different modes of transport. This is largely due to the interoperability issues that would otherwise arise for Intelligent Transport Systems and which could render those systems incompatible and potentially inoperable. It may therefore be helpful to adopt more technical specifications in the future, for instance in the form of additional data sharing obligations.

In terms of remuneration, a distinction should be made between situations where data must be provided to public authorities only and those where the data is to be shared to a wider community including private stakeholders. The former should typically be provided free of charge, while for the latter some kind of reasonable remuneration may be appropriate.

When data sharing obligations are imposed through legislation, policy makers can also introduce requirements relating to the conditions for access and re-use. This too can be an opportunity to achieve harmonisation, for instance by consistently referring to the conditions for access and reuse that are also imposed on public sector bodies in the Public Sector Information (PSI) Directive. The EU and other policy makers can support data sharing by:

- Legally requiring private companies to share the data generated by them;

- Introducing technical specifications for such data to be shared by private companies;
- Introducing harmonised requirements relating to the conditions for access and re-use of data to be shared by private companies.

Table 10: Results, barriers and impact of policy actions related to data sharing in agreements

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Data sharing obligations through public tendering	Increased access to data otherwise held and maintained by a limited number of private companies. Increased data standardisation through harmonised technical and access requirements.	Unwillingness on part of private companies to share data due to legal, commercial and technical challenges.	Lowered barrier to provide big data applications. Higher innovation expected.
Data sharing obligations through legislation	Increased access to data otherwise held and maintained by a limited number of private companies. Increased data standardisation through harmonised technical and access requirements.	Unwillingness on part of private companies to share data due to legal, commercial and technical challenges.	Lowered barrier to provide big data applications. Higher innovation expected.

Supporting industry sectors

A report by the OECD on data-driven innovation noted that there are an increasing number of mergers between big data companies over the past years.[11] Unfortunately, current evaluation frameworks at EU and national levels often lack the tools to appropriately address big data-related issues in the context of merger control. As a consequence, mergers of certain (small or large) data-rich companies may introduce barriers to competition that can hinder innovation in the EU internal market. The EU can support the different players in industry sectors, including the transport sector, by adopting appropriate merger control tools.

Policy action 13 Merger control

The concentrations that are notified to the European Commission increasingly involve companies active in the collection and processing of big data. Where merger control regimes have traditionally been based on market share and sales figures, consideration should now be given to whether there is a need to add "deal-size thresholds" with a view to catching high value acquisitions currently falling outside merger control due to small or insignificant turnover figures of the target.

Another specific issue that arises in the context of mergers between big data companies is the classification of data itself. Data is an asset, and more specifically an essential asset for the functioning of certain undertakings. The quality and quantity of the data they control as well as the intellectual property rights that they possess in order to make use of data in certain ways constitute a new parameter for business operations in the digital world. It is precisely these assets as opposed to financial turnover that entices larger undertakings to acquire smaller companies with a large data-asset collection.

The EU can support industry sector through the following actions in the context of mergers and acquisitions:

- Assess whether there is a need to add "deal-size thresholds" with a view to catching high value acquisitions currently falling outside merger control;
- Incorporate in the merger control framework other thresholds alongside financial ones, such as consumer thresholds, which would give a better view of the circumstances under which the merger is taking place;
- Introduce merger control criteria that take into account the kind of data is being acquired, how unique it is, whether it can be easily replicated and whether rivals can be shut out, i.e. criteria that focus on data quality rather than solely on data quantity;

11 OECD, *Data-Driven Innovation: Big Data for Growth and Well-Being*, (OECD Publishing 2015) 94

- Assessing whether competition law should be adapted to support data standardisation efforts.

Table 11: Results, barriers and impact of policy actions related to supporting industry sectors

ACTION	EXPECTED RESULT	BARRIERS TO OVERCOME	EXPECTED IMPACT
Merger Control	Stricter evaluation of data-rich mergers that could otherwise generate anti-competitive effects in the future.	Current merger control frameworks often lack the tools to appropriately address big data-related issues.	Broader access to high-quality data. Increased competition through elimination of barriers introduced by mergers of companies with high-value datasets.

Conclusion and future outlook

The transport sector has constantly collected and analysed large amounts of data, however, recent developments in the quantity, complexity and availability of data collected from and about transport, together with advances in information and communication technology, are presenting new opportunities to create more efficient and smarter transport systems for public and freight.

The use of big data in the transport sector is relevant for governments (e.g. traffic control, planning and modelling, route planning and congestion management), for the private sector (e.g. travel industry, route planning and logistics, and competitive advantages) and for individuals (e.g. route and travel planning). Big data has opened a wide spectrum of opportunities in the field of transport research. Several challenges constitute opportunities for the industry and scientists in the foreseeable future. The LeMO roadmap for big data in transport in Europe focuses on ensuring the crucial conditions for the development of a European big data in transport ecosystem based on the research outcomes of the project.

We have presented one of the key outcomes of the LeMO project: a European roadmap for big data in transport. This roadmap is divided in two sections. **The research roadmap focuses on identifying the research priorities at stake, and the policy roadmap provides a set of recommendations to build a policy framework which will ensure that European transport sector can make the most out of the data transition.**

The policy roadmap is more concerned with policies that will resolve economic and political challenges together with legal, social and ethical issues. This highlights technology innovations which are usually neutral until they become embedded in larger social systems. Consequently, the policy roadmap focuses on the digital single market, cross-border data arrangements, government services, public investment and resource management. The research roadmap focuses more on technological resources and capabilities. **The LeMO research and policy roadmap is oriented towards a range of different stakeholders who will be prepared to consider their contributions to these policy and research actions, including policymakers, civil society organisations, institutions, business and industry bodies and academics.**

This roadmap ends with a suggestion for the implementation of the actions. The results of one of the surveys provide a basis for this. **Some of the actions identified in the roadmap have a higher priority than others for certain reasons.** LeMO's goals can be achieved more quickly if research and policy focus on actions that are more of a priority than others, have a higher impact or are easier to implement. Figure 3 shows how the 24 actions score on three common assessment criteria, namely the urgency, impact and feasibility of an action[12]. In this context, *urgency* relates to the speed with which an action has to be implemented and indicates the time-frame within which the action should be carried out. *Impact* signifies the effectiveness of an action and the usefulness of an action in achieving the larger objective. *Feasibility* encompasses the political, financial, ethical, legal or technical practicality of implementing the action.

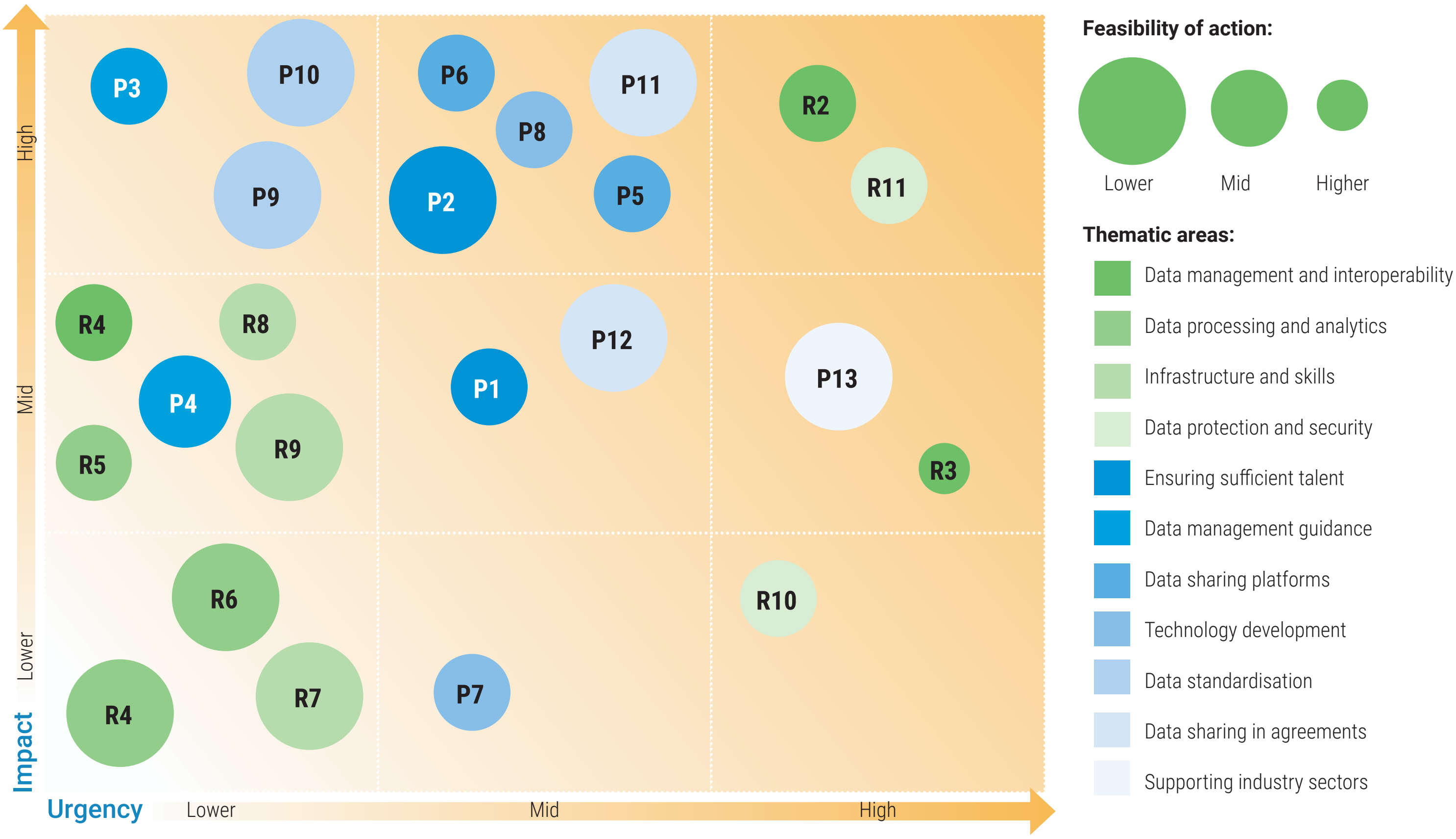
12 This figure is the result of the survey conducted among the advisory board and other data experts previously involved in *LeMO*.*In* total, 19 respondents completed the survey from the following fields: public sector (11); ICT sector (2); transport sector (1); Knowledge Institutes (4), Stakeholder organisation (1).

We do realise that the rating is a simplification of the complexity of actions. The rating should therefore be seen as recommendation to prioritize actions and as a source of inspiration for further research in this area. Nevertheless, on the basis of the rating, it is possible to identify certain high-priority actions. Research into enhancing cyber security (R11) and the development of techniques to improve data quality (R2) should receive high priority. With regards to the policy actions, the development of training and education programmes (P2), including data sharing obligations in public tendering (P11) and developing stricter merger control regimes (P13) are considered high priority policy actions. Actions scoring high on feasibility is an indication that these actions can be carried out as *Quick Wins*, such as the harmonisation of data-related legislation between EU Member States (P9) and the development of scalable and real-time data analysis algorithms (R6).

In conclusion, the proposed LeMO Research and Policy Roadmaps include **a set of actions from the research and policy point of view that governments, businesses and industries in Europe should implement in order to effectively take advantage of big data solution with the aim to provide added value for the society for a safe, secure, green, sustainable, accessible, cost-effective, inclusive and smart European transport model.**

Figure 3: Rating of the roadmap actions according to its urgency, impact and feasibility.

The numbers refer to the numbering used in Figure 2. A lower score means that the action scores lower relative to other actions while a higher score means that it scores above average.



References

LeMO deliverables

Jasper Tanis et al. Big data policies. D1.2 LeMO Project, Date: 31/05/2018.

Jayant Sangwan et al. Roadmap-Research and Policy Recommendations for the Efficient Utilisation of Big Data in the Transport Field (Version 1). D4.3 LeMO Project, Date: 30/06/2020.

Julien Debussche et al. Report on Legal Issues. D2.2 LeMO Project, Date: 31/10/2018.

Julien Debussche et al. Report on Ethical and Social Issues. D2.3 LeMO Project, Date: 31/08/2018.

Kim Hee et al. Big Data Methodologies, Tools and Infrastructures. D1.3 LeMO Project, Date: 31/07/2018

Marten Rosselli et al. Understanding and mapping big data in transport sector. D1.1 LeMO Project, Date: 31/05/2018.

Minsung Hong et al. Report on the characterization of the barriers and limitations. D4.1 LeMO Project, Date: 30/09/2019.

Minsung Hong et al. Horizontal Analysis and Socio-Economic Impacts. D4.2 LeMO Project, Date: 15/03/2020.

Otto Andersen et al. Report on trade-off from the use of big data in transport. D2.4 LeMO Project, Date: 30/11/2018.

Tharsis Teoh et al. Report on Economic and Political Issues. D2.1 LeMO Project, Date: 31/08/2018.

Tharsis Teoh et al. Case study reports on constructive findings on the prerequisites of successful big data implementation in the transport sector. D3.2 LeMO Project, Date: 30/06/2019

Literature

Aparicio, M., Oliveira, T., Bacao, F., & Painho, M. (2019). Gamification: A key determinant of massive open online course (MOOC) success. *Information & Management*, 56(1), 39-54.

Asensio, O., Alvarez, K., Dror, A., Wenzel, E., Hollauer, C., & Ha, S. (2020). Real-time data from mobile platforms to evaluate sustainable transportation infrastructure. *Nature Sustainability*, 1-9.

Axelrod, C. (2017). Cyber security in the age of autonomous vehicles, intelligent traffic controls and pervasive transportation networks. In 2017 IEEE Long Island Systems, *Applications and Technology Conference (LISAT)*, 1-6.

CAIRE, J. (2017). Human factors in cyber security for transportation systems. *WIT Transactions on the Built Environment,* 176, 405-414.

Carenini, A., Ugo, D. A., Stefanos, G., Pourhashem Kallehbasti, M. M., Rossi, M. G., & Riccardo, S. (2018). ST4RT–semantic transformations for rail transportation. *Transport Research Arena TRA 2018* (pp. 1-10).

Chalkiadakis, C., Iordanopoulos, P., Malin, F., Flachi, M., Giannini, M., Mitsakis, E., & Öörni, R. (2019). An Online Training Tool for Better Understanding the Operation and Significance of ITS. *6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)* (pp. 1-8). IEEE.

Chinrungrueng, J., Kovavisaruch, L. O., Narupiyakul, L., & Wongpatikaseree, K. (2018). Open Traffic Data Exchange and Collaborative Platform. In: *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)* (pp. 1-4). IEEE.

Ferdowsi, A., Challita, U., & Saad, W. (2019). Deep learning for reliable mobile edge analytics in intelligent transportation systems: An overview. *IEEE vehicular technology magazine*, 14(1), 62-70.

Ganzha, M., Paprzycki, M., Pawłowski, W., Szmeja, P., & Wasielewska, K. (2017). Semantic interoperability in the Internet of Things: An overview from the INTER-IoT perspective. *Journal of Network and Computer Applications*, 81, 111-124.

Jaisimha, A., Khan, S., Anisha, B. S., & Kumar, P. R. (2020). Smart Transportation: An Edge-Cloud Hybrid Computing Perspective. In: *Inventive Communication and Computational Technologies* (pp. 1263-1271). Springer, Singapore.

Katsumi, M., & Fox, M. (2018). Ontologies for transportation research: A survey. *Transportation Research Part C: Emerging Technologies*, 89, 53-82.

Li, Q., Chen, P., & Wang, R. (2019). Edge Computing for Intelligent Transportation System: A Review. In: *Cyberspace Data and Intelligence, and Cyber-Living, Syndrome, and Health* (pp. 130-137). Springer, Singapore.

Li, Y., Shu, H., & Deng, H. (2018). Privacy Protection of Transportation Big Data in Transportation: A Life-Cycle Perspective. *CICTP 2017: Transportation Reform and Change—Equity, Inclusiveness, Sharing, and Innovation* (pp. 346-355). Reston, VA: American Society of Civil Engineers.

Luo, T., Huang, J., Kanhere, S. S., Zhang, J., & Das, S. K. (2019). Improving IoT data quality in mobile crowd sensing: A cross validation approach. *IEEE Internet of Things Journal*, 6(3), 5651-5664.

Megler, V. M., Tufte, K., & Maier, D. (2016). *Improving data quality in intelligent transportation systems.* arXiv preprint arXiv:1602.03100.

Munir, M. S., Abedin, S. F., Kim, K., & Hong, C. S. (2019). Towards Edge Intelligence: Real-Time Driver Safety in Smart Transportation System. *Korea Computer Congress 2019* (KCC 2019) (pp. 26-28).

Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448.

Qin, Y., Luo, H., Zhao, F., Zhao, Z., & Jiang, M. (2018). A traffic pattern detection algorithm based on multimodal sensing. *International journal of distributed sensor networks*, 14(10), 1550147718807832.

Schoedon, A., Trapp, M., Hollburg, H., Gerber, D., & Döllner, J. (2019, September). Web-based Visualization of Transportation Networks for Mobility Analytics. *Proceedings of the 12th International Symposium on Visual Information Communication and Interaction* (pp. 1-5).

Seedah, D. P., Choubassi, C., & Leite, F. (2016). Ontology for querying heterogeneous data sources in freight transportation. *Journal of Computing in Civil Engineering*, 30(4), 04015069.

Stefanouli, M., & Economou, C. (2018, May). Data Protection in Smart Cities: Application of the EU GDPR. *The 4th Conference on Sustainable Urban Mobility* (pp. 748-755). Springer, Cham.

Sun, W., Bocchini, P., & Davison, B. D. (2020). Resilience metrics and measurement methods for transportation infrastructure: the state of the art. *Sustainable and Resilient Infrastructure*, 5(3), 168-199.

Sun, Y., & Song, H. (Eds.). (2017). *Secure and trustworthy transportation cyber-physical systems*. Springer Singapore.

Torre-Bastida, A., Del Ser, J., Laña, I., Ilardia, M., Bilbao, M. N., & Campos-Cordobés, S. (2018). Big Data for transportation and mobility: recent advances, trends and challenges. *IET Intelligent Transport Systems*, 12(8), 742-755.

Tucci, A. E. (2017). Cyber risks in the marine transportation system. In: *Cyber-Physical Security* (pp. 113-131). Springer, Cham.

Westerlund, M. (2018). *A study of EU data protection regulation and appropriate security for digital services and platforms*. Åbo Akademi University.

Yang, W., Sun, Y. E., Huang, H., Du, Y., Huang, D., Tao, T., & Luo, Y. (2020). Persistent transportation traffic volume estimation with differential privacy. *Journal of Ambient Intelligence and Humanized Computing*, 1-19.

Contact us

Rajendra Akerkar, LeMO Project Coordinator
Western Norway Research Institute - Sogndal, Norway
E-mail: rak@vestforsk.no
(t) +47 91685607

www.bigdata.vestforsk.no

Design inextremis.be / Photos AdobeStock