LEVERAGING BIG DATA FOR MANAGING TRANSPORT OPERATIONS

Deliverable 4.1 **Report on the characterization of the barriers and limitations**

Minsung Hong[1], Rajendra Akerkar[1], Tharsis Teoh[2], Timo Eichhorn[3], Julien Debussche[4]

Western Norway Research Institute[1],

Panteia B.V.[2],

Goethe-University Frankfurt[3]

Bird&Bird[4]

September 2019

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	30/09/2019 (M23)
Sent to coordinator	29/09/2019
No. of document	D4.1
Title	*Report on the characterization of the barriers and limitations*
Status & Version	*Final version*
Work Package	*4: Horizontal analysis and road-mapping*
Related Deliverables	*D4.2, D4.3 and D4.4*
Leading Partner	*WNRI*
Leading Authors	*Minsung Hong, WNRI* *Rajendra Akerkar, WNRI*
Contributors	*Tharsis Teoh, Panteia* *Timo Eichhorn, GUF* *Julien Debussche, B&B*
Reviewers	*Rémy Russotto, CORTE* *Marusa Benkic, CORTE* *Jayant Sangwan, CORTE*
Keywords	*big data, barriers and limitations, case study, technological, policy, legal, ethical, social, economic, political and environmental issues*

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 identifies and characterises barriers and limitations of the transport systems for exploiting big data opportunities. The scope of analysis in the document covers technological issues, as well as social, ethical, legal, political, organisational and institutional aspects of the systems. The characterisation of the barriers takes into account the extent to which they can constitute an absolute limitation for exploiting big data.

We extracted total 129 barriers (and limitations) from the various aspects and identified 54 from case study of LeMO, use case of NOESIS and pilot systems of TT. From the samples studied under the three projects, exciting findings were figured out as follows:

- It can be seen that technological and economic & political aspects are being considered and discussed more by the current industry.
- It has been seen that there are potential barriers and limitations in legal aspects that are many and of various types. It would have been spurred by GDPR regulations.
- Moreover, there are links between the barriers falling under different aspects. This further complicates the impact of these barriers on the use of big data in transport.
- Furthermore, it has been observed that there are interventions that can diminish the negative impacts of the barriers or convert them into opportunities. Sometimes, the barriers and opportunities also often mutually affect each other.

These observations will be further discussed and explored in the future deliverables in this work package. Especially, LeMO Task 4.2 will conduct a horizontal analysis to produce constructive findings and suggestions for successful big data implementation in the transport sector.

Table of contents

List of Figures

List of Tables

Glossary

Abbreviation	Expression
LeMO	Leveraging big data for Managing Transport Operations
NOESIS	Novel Decision Support tool for Evaluating Strategic big data investments in Transport and Intelligent Mobility Services
TT	Transforming Transport
KPI	Key Performance Indicator
N#	Categories of various factors by which NOESIS's use case have been evaluated
T#	Classifications of aspects that have been used to evaluate TT's pilot systems
CS#	Case studies of LeMO project
P#	Pilot systems of TT project
ERTMS	European Railway Traffic Management System
ETCS	European Train Control System
ATP	Automatic Train Protection system
GSM-R	Global System for Mobile Communications – Railway
AFC	Automated Fare Collection
APC	Automated Passenger Counter
GDPR	General Data Protection Regulation
NIS Directive	Directive on security of network and information systems
ENISA	European Union Agency for Network and Information Security
ITS	Intelligent Transport Systems
ROI	Return-on-Investment
WP	Work Package

1 Introduction

1.1 Abstract

This deliverable evaluates barriers and limitations of utilising big data in transport operations. In WP3, seven case studies conducted to create an inventory of opportunities, limitations and barriers in the utilisation of big data in transport. These were categorised under technological, economic & political, policy & legal and social & ethical aspects. These case studies were drawn from various themes related to different transport modes and sectors.

This deliverable presents a comprehensive evaluation of the effects and issues associated with the use of big data in ***LeMO case studies, as well as in the other relevant big data initiatives.*** The next task 4.2 will conduct horizontal analysis to diminish the negative impacts (limitations and barriers) and further to augment the positive effects (opportunities). This report and D4.2 will lead to identification of challenges, opportunities, limitations and other consequences of cross-disciplinary nature, and thus relevant to transport big data as a whole.

1.2 Purpose of the document

The purpose of this document is to identify and characterise barriers and limitations by analysing the outcomes of LeMO case studies (which cover all transport modes) and other relevant big data initiatives. The scope of the analysis will cover technical issues, as well as organisational and institutional aspects of the transportation systems. The characterization of the barriers will take into account the extent to which they constitute an absolute limitation for exploiting big data. This task will feed into other tasks under the work package relating to the horizontal analysis as well as development of research and policy roadmaps i.e. D4.3 and D4.4.

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 legislation.

2 Methodology

In this section, we describe the process of WP4 along with objectives of each task and relations between them. Also, definitions of terms, such as intervention, opportunity or barrier, are provided in order to achieve the WP4's objectives, which are as follows:

- To provide knowledge of barriers, and on the basis of this knowledge, provide recommendations for exploiting big data within the limitations that the barriers are creating.
- To undertake a horizontal analysis of case studies to identify constructive findings and suggestions that can be traced across contexts.
- To design research and policy roadmaps for efficient utilisation of big data in the transport field.
- To gain stakeholder consensus on the LeMO roadmap via a workshop.

Purpose of each task in WP4

Task 4.1: Barriers and limitations of the transportations systems for exploiting big data opportunities are identified in terms of the economic, legal, social, ethical and political issues, technologies and infrastructures as well as big data initiatives and policies. Moreover, we consider not only case studies of LeMO but also use cases of NOESIS[1] and pilot systems of TT[2] projects for consolidated analysis in subsequent tasks in WP4.

Task 4.2: This task will undertake a horizontal analysis of the case studies to produce constructive findings and suggestions on the prerequisites of successful big data implementation in the transport sector. We will also relate the findings of the case studies to various issues that we examine in work packages 1, 2, and 3, as well as in Task 4.1. As a result, we will examine the impacts of these findings and suggestions from a socio-economic point of view.

Task 4.3 and 4.4: Based on the findings and suggestions from the above tasks, LeMO will develop research and policy roadmaps for utilisation of big data to manage transport operations. Both roadmaps will provide a step by step process. The research roadmap will focus on knowledge surrounding economic, legal, social, ethical and political issues, as well as standards, interoperability and development of meta-data. The policy roadmap will focus on policies, investment, funding and infrastructure required to take full advantage of the opportunities surrounding big data. These roadmaps will result in D4.3 (i.e. the first version of the research and policy recommendations for the efficient utilisation of big data in the transport field).

Task 4.5: The D4.3 will be validated by the Advisory Board members and stakeholders in a validation workshop. All workshop participants will be provided with a draft copy of the two-

[1] NOESIS: https://noesis-project.eu/

[2] Transforming Transport: https://transformingtransport.eu/

part recommendations and be invited and encouraged to provide constructive feedback within breakout activities. The feedback will be used to revise and finalise D4.3 into D4.4.

Definition of terms in WP4

Barriers and limitations: Are obstacles or actions that inhibit the use of big data in the transport sector. The barriers can have different characterizations either technological; or policy and legal; or ethical and social; or environmental; or economic and political; or a mix of these. These barriers will be identified in the D4.1 by Task 4.1.

Opportunities: They occur when a product, activity or decision by an actor causes opportunities or benefits on the utilisation of big data in the transport sector. In this regard, suggestions for the positive effects can be provided as some ways to promote the opportunities and benefits and to foster the use of big data in the sector. It will be represented in D4.2.

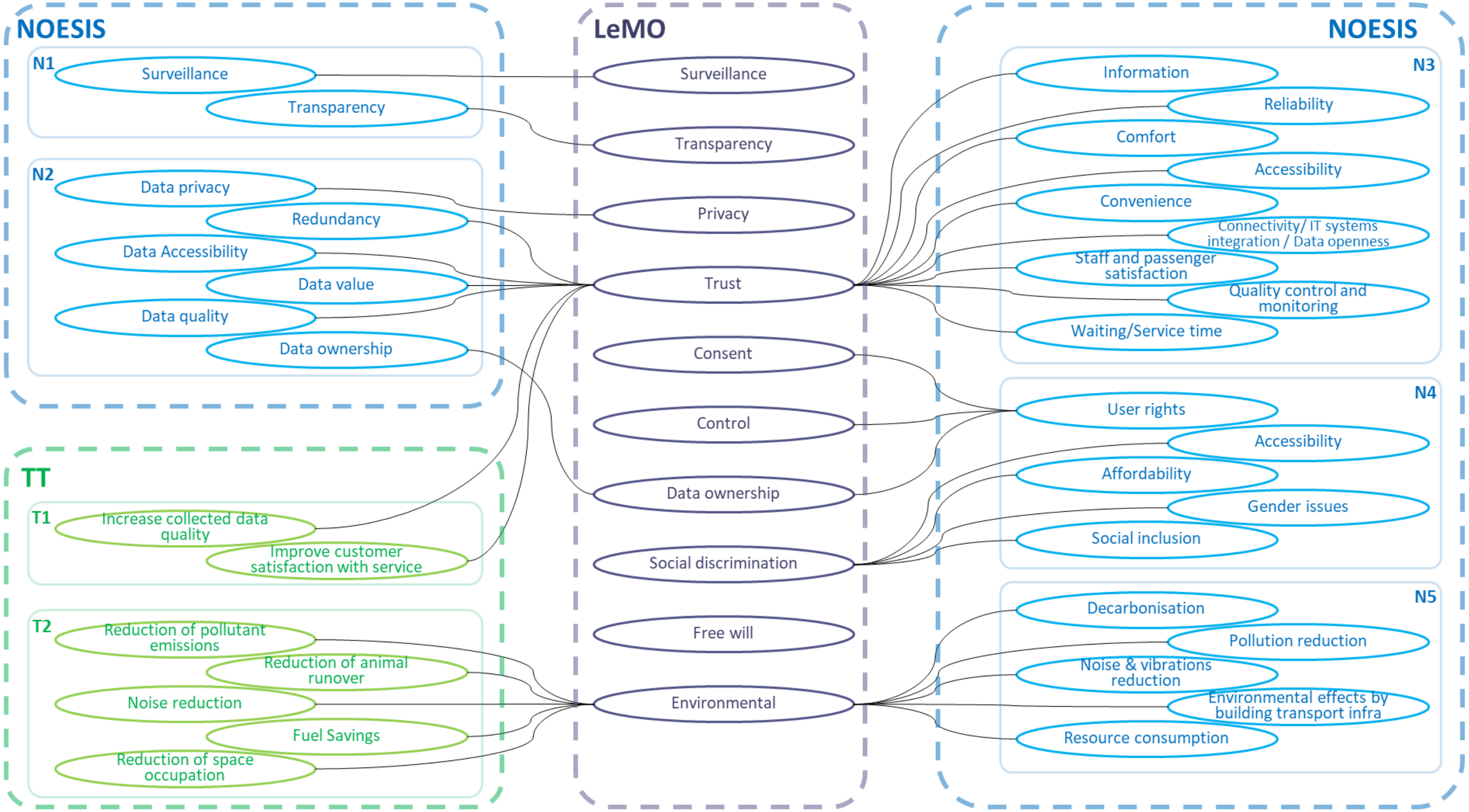
Interventions: These are steps that could be suggested to diminish the barriers/limitations in terms of the various issues or to convert the barriers into opportunities. While D4.2 will provide an outline of these interventions, they will be represented in greater detail in D4.3 and upon validation in D4.4.

Analysis in LeMO project is enriched by and compliments the work under other relevant EU Horizon 2020 initiatives such as Novel Decision Support tool for Evaluating Strategic big data investments in Transport and Intelligent Mobility Services (NOESIS) and Transforming Transport (TT).

LeMO preliminarily collected factors related to ethical and social issues, from both NOESIS and TT, and analysed relations between these factors, as shown in the following Figure 1. Note that several ethical and social issues extracted by LeMO project are covered by the factors identified under the two projects. For example, trust issue is related to quality of service (i.e., N3) from NOESIS project and operational efficiency (i.e., T1) from TT project, and environmental issue is overlapped with environment and health (i.e., N5) and environmental quality category (i.e., T2). However, some issues such as consent and control are partially considered by only the factor ‘user rights’ under NOESIS. Moreover, ‘free will’ is not dealt with under the two projects. On the other hand, several issues (i.e., trust, social discrimination and environment) have been considered in more detail under the two projects than in LeMO.

As we can see from above, analysing outcomes from different projects studying utilization of big data in transport is helpful, not only to gain a deeper understanding of identified issues but also to be complementary by apprehending missing issues. Therefore, LeMO will consider other big data initiatives, especially NOESIS and TT projects, as a cluster to inform the work under WP4.

Note that all identified barriers, opportunities and interventions will be presented by annotations in order to be referred to in the processes of WP4. In this deliverable 4.1, we use annotations, which are represented in Appendix A for barriers and limitations in terms of technological, policy, legal, ethical, social, environmental, economic and political issues. These will also be similarly used in subsequent outcomes.



N# indicates categories of various factors by which NOESIS's use case have been evaluated as follows: N1 - Safety and security, N2 - Data related factor, N3 - Quality of service, N4 - Social / Psychological aspects and N5 - Environment and Health. T# represents classifications of aspects that have been used to evaluate TT's pilot systems, (e.g., T1 - Operational Efficiency (OE) and T2 - Environmental Quality Category (EQ)). Note that we present only factors, which are related to social and ethical aspects, from the two projects. For other factors and its details, refer their websites.

Figure 1 Relations between ethical and social aspects from LeMO, NOESIS and TT projects

3 Technological barriers & limitations

To identify the technological barriers and limitations the LeMO case studies and the connatural EC projects TT and NOESIS were analysed. To start with, we identified the used data sources and compared these to the finding of deliverable D1.1. We also provide a mapping of the data sources into different transportation modes. To outline the solutions to overcome some of the barriers, we present a compendium of the data-driven technological activities we identified.

In deliverable D1.1 [1] data sources are divided into four different fields: Route-based data, Vehicle-based data, Traveller-based data and wide-area data.

Route-based data are collected by sensors at fixed locations of a path such as a highway or a train. The detectors are harnessed for monitoring intersection traffic, as well as detecting incidents, classifying vehicle and re-identifying vehicle.

Vehicle-based data are collected by mobile devices or in-vehicle GPSs. Whereas the route-based data are collected at a specific location, vehicle-based data are dynamic such as data of route choice, travel time estimation, and more. In particular, connected vehicle technologies enable vehicles to share data in real-time with other vehicles and the transport system.

Traveller-based data are collected by people. For example, traffic jams are inferred from one's location, and accidents are voluntarily reported by mobile device users.

Wide area/external data are collected by sensor networks to monitor traffic flow. Unmanned aircraft, space-based radar, and weather data fall into this type of data source.

In addition, traffic data collection techniques are reported to get an overview of how the numerous data is obtained. Table 1 lists all identified data sources in the LeMO case studies [2], NOESIS's use cases [3] and pilot systems of TT (Smart Highways [4], Connected Vehicles [5], Proactive Rail Infrastructure [6], Port Pilots Design [7], Airport Turnaround [8], Integrated Urban Mobility [9] and Dynamic Supply Chain Networks [10]).

The analysis shows that all identified fields of potential data sources are used in the investigated case studies. Also, for route and vehicle-based data, mainly different sensor technologies are used to provide the necessary data. Whereas traveller-based and wide-area/external data is collected by data providers and accessed by using their open data platforms or APIs. Sometimes this data is even purchased. We present each source with a classification to a transportation mode (Road, urban, rail, air and water mode) and the corresponding page in deliverable D3.2.

Table 1 Identified data sources

Sources of big data	Identified data sources	Collection techniques
Route-based data	• The European Railway Traffic Management System (ERTMS) is an industrial project in close cooperation with the European Union, railway stakeholders and the GSM-R industry, which enables the collection of data needed for predictive maintenance. ERTMS has two basic components: the European Train Control System (ETCS), which is an automatic train protection system (ATP), and GSM-R, a radio system for providing voice and data communication between the track and the train. [Rail mode, CS1] • Traffic data from sensors on roads/traffic flow [Road mode, CS1, CS3, CS4, CS5, CS7; NOESIS monitoring devices; TT Smart Highways, TT Integrated Urban Mobility] • Photos and videos from road CCTV cameras [Road mode, CS2, CS4, CS7] • Parking/mooring information [Road, water mode, CS2, CS5; TT Integrated Urban Mobility] • Environmental data on the condition/state of the road [Road mode, p. CS2, CS4, CS7] • Events and traffic incidents [Road mode, CS2, CS4, NOESIS Physical events; TT Smart Highways, TT Integrated Urban Mobility] • Data on the status of bridges whether a bridge is open or closed, a sensor is installed on the bridge or processes in the industrial control system of the bridge are read out. [Water mode, CS5] • Emergency services data and public transport data [Water/Road mode, CS5] • Road and waterway maps [Water/Road mode, CS5, CS2, CS7, NOESIS digital maps] • Local traffic management system. [Road mode, CS5; TT Smart Highways; TT Port Pilot Design, TT Integrated Urban Mobility] • Real time information of toll stations (queue length, toll events, etc.) [TT Smart Highways] • Transit data/Smart Cards: Smart card Automated Fare Collection [AFC] and Automated Passenger Counter [APC] are technologies for collecting public transit data to both describe the spatial-temporal patterns of passengers' behavioural and evaluate transit facilities. [Urban mode, CS7] • Arrival and departure flights information according to flight schedule at the airport [Air mode, TT D8.1]	• Sensor data of Acoustic sensors, Microwave radar, Lidar and active infrared sensors. • CCTV and traffic cameras • Near field transponders on ships • Sensors to detect open or closed bridges • Transit data/Smart Cards • Digital maps (e.g. Google Maps, Apple Maps, Bing, etc.) • Information about the synchronisation of traffic lights • Magnetic / Pneumatic loops

Vehicle-based data	• Internet of Things - accessible information about the status of sensors built-in vehicles [multimodal, CS1, CS4 (weather information); TT Connected Vehicles, TT Proactive Rail Infrastructure] • Location specific information about the vehicle [multimodal, CS2, CS7, NOESIS Location data; TT Integrated Urban Mobility] • Baggage checked by passengers [Air mode, TT Airport Turnaround] • Historical maintenance data [Rail mode, TT Proactive Rail Infrastructure] • System for providing information about the ship to other ships and to coastal authorities automatically [Water Mode, TT Port Pilot Design]	• Sensor data of: ○ Process data, Diagnostic messages, Log messages [Rail mode, p.9] • GPS data of the vehicles • Screening of checked-in baggage in an airport • Wi-Fi and Bluetooth connections • Active and passive RFID tags
Traveller-based data	• Location-Based Social Networks [Multimodal, CS7; NOESIS Social Media; TT Integrated Urban Mobility] • Purchased data from mobile phone providers [Multimodal, CS7] • Non-personal data related to passengers [Air mode, TT D8.1] • Boarding pass read at the security/screening entries [Air mode, TT D8.1] • Location at the airport of mobile phones connected to the airport WiFi [Air mode, TT D8.1]	• Social network data • Purchased location data of mobile phone providers
Wide area/external data	• Historical, current and forecast of weather data on roads/waterway [Road, water mode, CS2, CS4 (crowdsourcing mPING, climate.org published by UC Berkeley), CS5, CS7; NOESIS environmental; TT Smart Highways, TT Integrated Urban Mobility] • Schedule for public transports [Urban mode, CS2; TT Proactive Rail Infrastructure] • Historical accident data on a broad scale across the US by bringing data from different states. [Road mode, CS4] • Data from different bike solution providers [Road mode, CS6] • Flight Plans emitted by the airlines/pilots and received from Eurocontrol [Air mode, TT D8.1] • Consumption and transaction data (e.g. data coming from Oyster card, credit card spend data, fares, petrol prices, etc.) [NOESIS Consumption and transaction data; TT Dynamic Supply Chain Networks] • Real time travel times and congestion level provided by navigation service [TT Smart Highways] • Satellite images to detect not only the current state of a location but also its changes over time [TT Connected Vehicles]	• Open data platforms and APIs • Commercial very-high-resolution satellite images

Table 2 provides a mapping of which transportation mode uses what kind of different data based on the analysed LeMO case studies [2]. The information about the air mode was gained out of the TT project [3].

Table 2 Mapping of field of data source to transportation mode

		Field of data source			
		Route-based data	Vehicle-based data	Traveller-based data	Wide area/external data
Transportation Mode	Road Mode	X	X	X	X
	Urban Mode	X	X	X	X
	Rail Mode	X	X	X	X
	Air Mode	X	X	X	X
	Water Mode	X	X		X

The mapping shows that nearly every transportation mode use all of the different classes of the existing data sources. Only the water mode has not made use of traveller-based data, but this was expected as the investigated Case study and Pilots were focussing only on freight transportation services.

By analysing the different case studies, pilots and deliverables of the mentioned projects, the following technical barriers and limitations affecting data were identified.

Table 3 Barriers and limitations: Data resource

Code	Description	Source(s)
LIM-TM-DR-1	**Data sources** The need for more data is documented in many of the case studies. The challenge is to collect data due to the fragmentation of data sources (which can be historical, user/vehicle provided or real-time) and multiplicity of data formats.	Case study 6, Case study 2, Case study 4, Case study 6; NOESIS D2.1; LeMO D1.1 Analyzation of data sources: TT D7.3 [11]
LIM-TM-DR-2	**Data quality** At the same time, ensuring sufficient data quality is mentioned in nearly all of the case studies as a current issue.	Case study 2, Case study 4, Case study 5, LeMO D1.1; NOESIS D2.1

LIM-TM-DR-3	**Data exchange** A lack of standardisation in sharing and handling of already existing data was mentioned by many of the contributing partners. This includes data discovery and its availability (there are many open data platforms), data integration of the various data and combining/integrating the different data sources without a unified standardisation.	Data discovery: Case study 2 Data integration: Case study 3 Lack of standardisation: Case study 1, Case study 2, Case study 3, Case study 7, LeMO D1.1

In deliverable D1.3 [12], we described the heterogeneity of big data Technologies in the Hadoop Ecosystem. It follows the classification proposed by Ivanov [13] and made an intuitive representation of a big data platform with the concept of the heterogeneity paradigm. The heterogeneity paradigm is a concept to understand the core components of the Hadoop Ecosystem and investigates the interconnection between its components. It also maps each component into different layers to provide a comprehensive overview of the Hadoop Ecosystem. The concept consists of four layers, which are also called levels: hardware, management, platform, and application. Below, we provide a brief description of each of the levels, but for a better understanding, we recommend reading deliverable D1.3.

The **hardware layer** represents the server components of the system and the fact that they can vary in storage, memory and processor type and size. Using the right hardware modules for a particular application can be crucial for obtaining the best price-performance ratio. The **management layer** deals with the system resource management and offers services to the applications running on the upper layers. It is responsible for the management and optimal allocation and usage of the underlying hardware components. The **platform layer** represents the main storage and processing services that a big data platform provides. In the last years, Apache Hadoop with its two main components HDFS (data storage) and YARN (processing and resource allocation) has become the de-facto platform for big data. Finally, the **application layer** is hosting the great variety of big data applications (data retrieval, aggregation, and processing including data mining and analytics, machine learning, etc.) running on top of the services provided by the lower layers.

Table 4 classifies the case studies of LeMO conducted in work package 3 into the different layers of the big data heterogeneity concept. It is conspicuous that only one case study (i.e. case study 1) is providing some additional service as they offer a cloud-based big data platform. All the other observed companies make use of the existing hardware, management, and platform technologies. On top of this application covering their individual requirements are developed and deployed.

Table 4 Classification of case studies to technology layers

Case study	Hardware	Management	Platform	Application
1 - Railway transport			X	X
2 - Open data and the transport sector				X
3 - Real-time traffic management				X
4 - Logistics and consumer preferences				X
5 - Smart inland shipping				X
6 - Optimised transport & improved customer service				X
7 - Big data and intelligent transport systems				X

Table 5 lists technical barriers/limitations mainly dealing with a certain **complexity** mentioned by the case study interviewees regarding setting up the big data architecture and collecting and exploiting the necessary data:

Table 5 Barriers and limitations: Data complexity

Code	Description	Source(s)
LIM-TM-DC-1	**Choosing the right architecture** Beginning with the complexity of building up a big data infrastructure (on-premise, hybrid, or cloud) by choosing and configuring the corresponding technologies and tools. This also includes data analytics, Machine Learning, Deep Learning and Artificial Intelligence algorithms.	Case study 5, Case study 7
LIM-TM-DC-2	**Data collection and processing** Effective data collection and processing as well as converging different types of information into a single traffic condition representation is seen as very difficult. This also includes real-time data collection and processing.	Data collection and processing: Case study 7, LeMO D1.1 Converging information: Case study 3
LIM-TM-DC-3	**Exploiting data** Exploiting all relevant traffic data is complex or impossible. How to exploit big data issues (data sources of many different systems) in a vehicular environment, once the current models and	Case study 3, LeMO D1.1

	algorithms used in big data are physically and logically decentralized, but virtually centralized. Also, the latency of data analytics needs to be taken into account.	

Also, some **limitations of the used systems** were mentioned with respect to the challenges of using big data technologies and the huge amount of data:

Table 6 Barriers and limitations: Limitations of the used systems

Code	Description	Source(s)
LIM-TM-LS-1	**Limitations of infrastructure** 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].	Case study 3
LIM-TM-LS-2	**Limitation traffic management system** It becomes necessary to rethink and redesign advanced traffic management systems to accommodate such volumes of captured data. The BGV data service, for example, had to be constructed from scratch, and all participants needed the technology and software to send data to the processing data service. In general, a lack of the existing model structures for current travel behaviour analysis, as well as the lack of scalability in the models were identified.	Redesign TMS: Case study 3 BGV data service: Case study 5 Models: D1.1

In addition to the above, the technological barriers and limitation discussed below were mentioned and/or identified in the seven LeMO case studies, in LeMO deliverable D1.1, in NOESIS deliverables 2.1 [14] and 2.2 [3] and the TT pilots. They focus on issues, which are either (1) caused by or (2) affect technical relevance. These limitations can be classified into two major groups: Those caused by technical restrictions OR Those affecting the technical solution of big data in the area of transportation.

It must be noted that these limitations/barriers may not strictly qualify as technical, since technology is not the root cause of these barrier, but they still affect the technical aspects. It must also be kept in mind that the solutions to these barriers may not be achieved by applying technological changes alone. These limitations also overlap with limitations identified under policy and legal; or ethical and social; or environmental; or economic and political groups.

Table 7 Barriers and limitations: Affecting technical solutions

Code	Description	Source(s)
LIM-TM-AT-1	Lack of **skilled workers** / expert knowledge.	Case study 1, Case study 3, Case study 4, Case study 5; LeMO D1.1
LIM-TM-AT-2	Large **investments costs** by applying big data or predictive maintenance.	Case study 1
LIM-TM-AT-3	**Security, Privacy and Trust** o Data flow controls related to roles and **data ownership**, making the collection and publication of data more complex. o Addressing public concerns over **personal data privacy** – identified as one of the key long-term threats to creating successful crowd-sourced products and services. CS4 provider sees the main risk in terms of **confidence**, i.e. will the end-user be confident about the data. [...] Therefore, the main challenge is to ensure that the end-users are confident about the model that is presented to them.	Data ownership: Case study 2, LeMO D1.1 Personal data privacy: Case study 3, LeMO D1.1; NOESIS D2.1 Confidence: Case study 4 Confidential driver data: Case study 4 Cyber security: NOESIS D2.1
LIM-TM-AT-4	Focusing on the **heterogeneity of individual** travelling is another challenge. Knowledge of users' travel behaviour is not yet mature and studying and modelling user's acceptance factors and travel-related choices represent an urgent area for further research.	Case study 7, LeMO D1.1
LIM-TM-AT-5	Avoiding **exclusion of certain societal groups** from having access to MOD services, like elderly people or people with no or less digital knowledge.	Case study 7
LIM-TM-AT-6	**Multiplicity of applications** is a big hurdle from customer's perspective. Given that there are several options for mobility [such as bicycles, scooters, shared cars, taxi and public transport] each with its own individual application, end-users have to install and navigate through several applications on their phone. This fragmentation and absence of a single	Case study 6

	"one-stop shop" application can be off-putting and prevent the expansion of user base.	
LIM-TM-AT-7	**Lack of business models** for efficient data exploitation.	NOESIS D2.1

We assess that all the mentioned limitations can be handled by corresponding research work and/or initiatives of private and public companies.

4 Policy and legal barriers & limitations

4.1 Policy issues

Big data applications in the transport sector have achieved national and EU-level interest as a driver for future economic growth and at the same time a source of concern, notably in terms of negative socio-economic impacts. In the context of Deliverable D1.2 entitled "big data Policies", it is demonstrated that current policies implemented in the EU, its Member States and internationally, support or restrict the (re-) use, linking of and sharing of data, notably in the context of big data and in the transport sector. It further illustrates in selected examples of transport-related private companies, the types of private sector policies that have been adopted or promoted.

Deliverable D1.2 has shown that there are not any distinct big data policies. However, there exist some policies aimed at protecting the privacy of citizens through restrictions of personal data processing activities, but also others encouraging data sharing among private and public sector organisations. Some other initiatives further aim at developing policies that support the digitalisation of the transport sector. Some of the key areas of policy in the transport sector are for instance the implementation of Intelligent Transport Systems, the increased Open Data policies, Automated Driving, and Smart Mobility. Preceding and in light of these developments, the private sector has also moved ahead to incorporate the use of big data techniques into their own business models as processes or product innovations.

On such basis, the table below list three core barriers and limitations related to public and private policies based on the preliminary results included in D1.2. As for the public policies, additional barriers and limitations can be found under Section 4.2 related to the legal issues.

Table 8 Barriers and limitations: Public and private policies

Code	Description	Source(s)
LIM-POL-PU-1	Although there is a general tendency towards data openness and data sharing in public policies across the EU, including in the transport sector, there are still discrepancies between the different Member States and local approaches remain. Such situation hinders a true coordinated EU-wide approach to big data.	New (inspired by LeMO D1.2)
LIM-POL-PU-2	In terms of public policies, a fundamental clash can be observed between the will to share as much data as possible on the one hand and the will to protect privacy on the other. This requires organisations to strike a difficult balance between data sharing/data openness and the protection of privacy.	New (inspired by LeMO D1.2)

LIM-POL-PR-2	Depending on many factors (e.g. an organisation's quest for profit, its size, type, legacy, data-dependency, etc.), organisations adopt very diverging internal/private data sharing policies and practices. This makes it difficult (and time-consuming) to anticipate how different organisations will engage in data sharing.	New (inspired by LeMO D1.2)

4.2 *Legal issues*

The legal barriers and limitations examined in the sections below derive from the analysis of the legal aspects included in deliverable D2.2 "Report on Legal Issues" of the LeMO project, as well as the LeMO case studies of Work Package 3. In addition, the findings of other EU projects were examined, such as in particular the result of EC projects TT and NOESIS.

In the below, we shortly introduce the following core 13 legal issues identified in the context of Deliverable D2.2:

Figure 2 Overview of 13 legal issues identified in Deliverable D2.2

On such basis, we provide for each of the 13 legal topics a table listing the barriers and limitations based on the preliminary results included in D2.2 but also on additional findings. The source of each barrier and limitation is indicated in the last column of the tables.

4.2.1 Privacy and data protection

Certain principles and requirements related to privacy and data protection can be difficult to fit with some of the main characteristics of big data analytics. In this respect, Deliverable D2.2 demonstrated that finding a balance between the various interests at stake is of paramount importance. In light hereof, 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.

A full analysis of the issues related to privacy and data protection is included in Deliverable D2.2, Section 3.1, pp. 9-63.

Table 9 Barriers and limitations: Privacy and data protection

Code	Description	Source(s)
LIM-LEG-DP-1	The GDPR is drafted in a technology-neutral manner, which means it can at times be difficult to apply to specific cases of personal data processing, such as in a big data analytics context. The legal assessment for privacy-compliant of big data analytics requires a case-by-case analysis of the specificities of the processing activities and the factual background, requiring a concrete application of the general rules.	New
LIM-LEG-DP-2	Certain key concepts (e.g. "personal data" and "processing") are defined and interpreted in such a broad way that the strict rules of the GDPR would apply, to such extent that this would prohibit, hinder, require additional investment or slow down the development of certain technologies or processing activities.	LeMO D2.2 LeMO Case Study 2 LeMO Case Study 5 LeMO Case Study 6
LIM-LEG-DP-3	Merely technically collected data could, due to the transformational impact of big data analytics, become personal data or even sensitive data and thus trigger the application of privacy and data protection laws.	LeMO D2.2 NOESIS D4.1 LeMO Case Study 2
LIM-LEG-DP-4	Some interpretations provided by certain authorities are conservative, too restrictive, and/or simplistic to such extent that it would prohibit, hinder, require additional investment or slow down the development of certain	LeMO D2.2

	technologies or processing activities (e.g. in relation to the further processing of personal data, the extent of data portability, etc.).	
LIM-LEG-DP-5	There is a lack of agreement on interpretation on certain issues between various supervisory authorities across the European Union.	New
LIM-LEG-DP-6	The distinction between "controller" and "processor", taking into account the concepts of joint-controllership, controllers in common and sub-processors, is complex in a big data context. Hence, identifying the role of each actor intervening in a big data context might prove to be difficult.	LeMO D2.2
LIM-LEG-DP-7	Technologies such as big data analytics challenge certain core assumptions of the EU data protection law, such as data minimisation and purpose limitation.	LeMO D2.2 NOESIS D4.1
LIM-LEG-DP-8	The processing purpose and the compatibility with that purpose in the event of any further processing of personal data can be particularly difficult to establish in light of the characteristics of big data (i.e. merging different datasets regardless of their purpose and analysing them to develop new business models).	NOESIS D4.1
LIM-LEG-DP-9	Big data analytics often uses complex algorithms that are difficult to understand by data subjects. Such difficulty poses issues in relation to the transparency principle and requirements.	LeMO D2.2 NOESIS D4.1
LIM-LEG-DP-10	Big data applications must be designed to meet the rights of data subjects (e.g. right to access or right to erasure), which may prove to be particularly difficult.	NOESIS D4.1
LIM-LEG-DP-11	The grounds permitting the processing of personal data, exhaustively listed in the GDPR, will generally be difficult to apply in a big data context.	LeMO D2.2 NOESIS D4.1
LIM-LEG-DP-12	The various conditions of consent are stringent and may be particularly difficult to meet in many instances. Therefore, relying on consent can be particularly difficult or may prove to be unpractical or even impossible in a big data context, especially in its more complex applications.	LeMO D2.2 NOESIS D4.1

LIM-LEG-DP-13	The stakeholders active in the context of disruptive technologies, including big data analytics, are required to conduct one or more data protection impact assessments (DPIA's) prior to the data processing, and continuously update such assessments. Notably, with self-learning applications, the issue arises whether a risk analysis in the context of a DPIA can be performed at all when these applications continue to self-develop.	LeMO D2.2 NOESIS D4.1
LIM-LEG-DP-14	The requirements related to the international transfer of personal data must be taken into account in order to determine the adequate solution to permit such international flows. This requires an extensive mapping of all international data flows, which can be rather challenging in a data-rich environment involving numerous actors and relying on cloud computing technologies.	LeMO D2.2

4.2.2 (Cyber-)Security and breach-related obligations

Considering the increasingly devastating impact that cyber-threats and attacks may have on society, issues related to cyber-security have become more and more important in recent years. The requirement to put in place security measures is imposed in various legislations at EU and national level, including key instruments like the General Data Protection Regulation (GDPR) and Directive 2016/1148 on security of network and information systems (the NIS Directive). Such legislations however remain rather general and vague as to which specific measures are deemed appropriate. In order to comply with the relevant requirements, organisations generally need to rely on security experts and take into account the evolving guidance documents published by authorities such as ENISA (the European Union Agency for Network and Information Security). Also, relying on certification mechanisms, seals, marks and codes of conduct will enable companies to comply with their legal obligations in terms of security and demonstrate their compliance.

A full analysis of the issues related to (cyber-)security is included in Deliverable D2.2, Section 3.2, pp. 64-82.

In recent years, the EU has made significant progress in terms of cybersecurity and related incident notification requirements, with notable developments including the Cyber Security Strategy and the NIS Directive. It follows that organisations facing a security incident may need to notify such incident to one or more national competent authorities. The requirement to inform authorities will however depend on certain criteria laid down in the applicable legislations, as clarified by the guidance documents published at EU and national level. Accordingly, the various actors of the data value chain need to implement measures, procedures and policies in order to abide by the strict notification requirements and be prepared to provide the necessary information to the competent authorities, all within the imposed deadlines.

A full analysis of the issues related to breach-related obligations is included in Deliverable D2.2, Section 3.3, pp. 83-92.

Table 10 Barriers and limitations: (Cyber-) security and breach-related obligations

Code	Description	Source(s)
LIM-LEG-CBO-1	Big data applications carry a higher risk of attacks due to the vast amount of data being processed. Due to the multitude of data, such a system is a rewarding target for malicious actors.	NOESIS D4.1 LeMO Case Study 3
LIM-LEG-CBO-2	Due to the large amount of data processed through big data applications, any loss of data is a considerable one.	NOESIS D4.1 LeMO Case Study 3
LIM-LEG-CBO-3	Big data analytics may pose specific security issues such as in relation to access control and authentication, secure data management, and source validation and filtering.	LeMO D2.2
LIM-LEG-CBO-4	Legal instruments covering security aspects generally impose general requirements, such as in particular putting in place "appropriate technical and organisational measures to ensure a level of security appropriate to the risk". Especially in a big data context, it will be difficult to assess what level of security is "appropriate to the risk", as it will be difficult to assess the risk to begin with.	New
LIM-LEG-CBO-5	Determining the appropriate security measures for big data application requires taking into account several factors, such as state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity (for the rights and freedoms of natural persons). Such an assessment may prove to be particularly difficult in practice.	New
LIM-LEG-CBO-6	Threats to security are an ever-evolving issue which requires keeping up-to-date risk assessments and	LeMO D2.2

	upgrading security measures. This may prove difficult in the event of continuous big data analytics.	
LIM-LEG-CBO-7	Certain legislations, such as the NIS Directive, impose requirements that may impact a multitude of actors of the data value chain, including those indirectly impacted due to flow-down obligations.	LeMO D2.2
LIM-LEG-CBO-8	Given the nature of certain legal instruments, such as the NIS Directive (i.e. minimal harmonisation Directive), discrepancies across the EU may exist in terms of what type of organisations fall within the scope of, and therefore need to comply with, the applicable security and breach-related obligations.	New
LIM-LEG-CBO-9	Given the nature of certain legal instruments, such as the NIS Directive (i.e. minimal harmonisation Directive), discrepancies across the EU may exist in terms of measures that must be put in place in order to adequately and appropriately secure data and to notify incidents or data breaches. It may therefore be particularly difficult to comply with diverging requirements in large and cross-border big data analytics applications.	LeMO D2.2
LIM-LEG-CBO-10	The interaction with authorities may prove difficult in many circumstances given the possible existence of multiple competent authorities in each country, depending notably on the applicable legislation. It follows that the legal assessment may be complex in case of big data projects involving multi-modal and cross-border transport services including numerous actors.	LeMO D2.2

4.2.3 Anonymisation and pseudonymisation

Anonymisation and pseudonymisation techniques have an impact on a personal data protection context, but their use is also a way to protect non-personal data. Anonymisation and pseudonymisation techniques generally provide fertile ground for opportunities with respect to big data applications. Nevertheless, account must be taken of the challenges that may arise in this respect. Most importantly, a balance will need to be struck between, on the one hand, the aspired level of anonymisation (and its legal consequences) and, on the other hand, the desired level of predictability and utility of the big data analytics.

A full analysis of the issues related to anonymisation and pseudonymisation is included in Deliverable D2.2, Section 3.4, pp. 93-113.

Table 11 Barriers and limitations: Anonymization and pseudonymisation

Code	Description	Source(s)
LIM-LEG-AP-1	As the definition of personal data is constantly evolving, anonymisation techniques should also continuously evolve and become increasingly robust in order to achieve irreversible anonymisation (where desired).	LeMO D2.2
LIM-LEG-AP-2	A recent study has demonstrated that currently no adequately robust anonymisation techniques exist. More particularly, the study suggests that even heavily sampled anonymised datasets are unlikely to satisfy the modern standards for anonymisation set forth by data protection.[3]	New
LIM-LEG-AP-3	Even if it is not possible today to identify data subjects from an anonymous dataset, it may become possible in the future by using new and more sophisticated analytical methods. Accordingly, if one decides to use only anonymous data in a big data application, it shall ensure that the data remains anonymous throughout the life of that application.	NOESIS D4.1
LIM-LEG-AP-4	The data protection legislation will remain applicable if, in spite of the anonymisation techniques used, the data subject can still be identified. In such event, all data protection principles and obligations must be respected by the data controller as well as the data processor when processing the personal data.	LeMO D2.2
LIM-LEG-AP-5	The combination of (anonymous) data processed in a big data application increases the probability that data will become personal again, and thus triggering the application of the data protection legislation and the related requirements. It follows that the applied anonymisation techniques should be reassessed and adjusted when necessary.	NOESIS D4.1
LIM-LEG-AP-6	A too far-reaching anonymisation of data may limit predictability in the big data analytics.	LeMO D2.2
LIM-LEG-AP-7	The available guidance in relation to anonymisation and pseudonymisation in a personal data context is outdated and	New

[3] Luc Rocher, Julien M. Hendrickx and Yves-Alexandre de Montjoye, 'Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models' (2019) 10(3069) Nature Communications <https://doi.org/10.1038/s41467-019-10933-3> accessed 12 September 2019

	has not been endorsed or renewed by the European Data Protection Board.	

4.2.4 Supply of digital content and services

The EU institutions recently adopted Digital Content Directive (EU) 2019/770 which introduces a high level of protection for consumers paying for a service but also for those providing (personal) data in exchange for such service. Such an instrument is particularly relevant to assess in light of the possible provision of (personal) data by a consumer in order to receive digital content. It is notably interesting to examine the interaction of such legal instrument with the applicable data protection legislation, and in particular the GDPR. A more in-depth assessment of the phenomenon allows concluding that legalising this economic reality generates practical and legal concerns.

A full analysis of the issues related to the supply of digital content and services is included in Deliverable D2.2, Section 3.5, pp. 114-121.

Table 12 Barriers and limitations: Supply of digital content and services

Code	Description	Source(s)
LIM-LEG-DIG-1	It is likely that the data retrieval obligations upon termination of the digital content provision will be difficult to comply with from a technical perspective in a big data context (e.g. due to data isolation, anonymisation and pseudonymisation). In addition, it may prove difficult or even impossible to return the data to the user without collecting more data than currently collected. Such obligations are also likely to increase costs and lead to administrative burdens on the industries supplying digital content in the transport sector.	New (inspired by LeMO D2.2)
LIM-LEG-DIG-2	The data retrieval obligations upon termination of the digital content provision may have a negative impact on the remaining users' experience, given that the data provided or generated by the users accessing the digital content may be indispensable for the product or service to function for those other users. In some cases, this could go so far as to render certain current content and services inoperable.	New
LIM-LEG-DIG-3	Data is often used for a wide range of commercial purposes, which may involve "indirect monetisation", such as security or improvement of customer experience. The new legislative framework however fails to address the variety and the	New

	specificity of data uses. The question arises when exactly a trader has supplied or has undertaken to supply digital content and the consumer provided or has undertaken to provide personal data, within the meaning of the Digital Content Directive.	
LIM-LEG-DIG-4	Certain aspects of a digital content provision are not regulated by the Directive or are left to national legislation, and in particular aspects related to the trader's rights. This is likely to create legal uncertainty and/or discrepancies across the EU.	New
LIM-LEG-DIG-5	The interplay between the rules enshrined in the Directive and the strict requirements related to personal data as regulated by the GDPR is unclear and is likely to lead to legal uncertainty.	New (inspired by LeMO D2.2)
LIM-LEG-DIG-6	The obligations concerning data may make some current services inoperable. Some companies may also start to charge for digital content services that are currently free. On a wider scale the ecosystem of innovative services in the field of transport could be jeopardised. Many start-ups and small companies do indeed rely on free digital content for their business model.	LeMO D2.2

4.2.5 Free flow of data

Free flow of data represents an ideal scenario in which no (legal) barriers to cross-border data flows remain. Efforts have been taken at EU level with the adoption of the Regulation on the free flow of non-personal data. However, a number of uncertainties remain, including a difficult interaction with the GDPR. Still, the Regulation remains an important step in the elimination of restrictions to cross-border data flows and their negative impact on business. Companies expect cost reductions to be the main benefit of eliminating data localisation requirements. Furthermore, start-ups in the European transport sector and in the EU in general increasingly rely on competitive cloud services for their products or services. Prohibiting localisation restrictions would therefore increase the competitiveness of the EU cloud services market. This in turn could allow start-ups to go to market quicker, to increase their pace of innovation and would also support scalability and achieve economies of scale.

A full analysis of the issues related to the free flow of data is included in Deliverable D2.2, Section 3.6, pp. 122-134.

Table 13 Barriers and limitations: Free flow of data

Code	Description	Source(s)
LIM-LEG-FF-1	The Free Flow Regulation will only apply to non-personal data. Such scope of application entails uncertainty as "personal data" is broadly defined. Also, even when data is non-personal it can become personal by merging it with other (non-)personal datasets, as is common in a big data context.	LeMO D2.2 NOESIS D4.1
LIM-LEG-FF-2	Following recent guidance, GDPR will apply without limitation to mixed datasets composed of personal and non-personal data that are inextricably linked, which will very often be the case for big datasets in the transport sector. It is unclear how such a situation will be resolved in practice.	New
LIM-LEG-FF-3	The Free Flow Regulation does not regulate the situation in which cross-border access to non-personal data by competent authorities is prohibited by the Member State in which the data is located.	New (inspired by LeMO D2.2)
LIM-LEG-FF-4	No safeguards (e.g. for third parties' IP rights or to protect the commercial value of trade secrets) are established concerning cross-border access by competent authorities to non-personal data.	LeMO D2.2

4.2.6 Intellectual property rights

Several intellectual property rights could be particularly relevant when examining the protection of data and in particular to what extent such protection mechanisms can apply to (big) data. More specifically, the aspects related to copyright, database rights and trade secrets are particularly relevant. In this respect, it cannot be excluded that different actors in the big data analytics lifecycle will try to claim intellectual property rights or protection under trade secrets in (parts) of datasets intended to be used. These actors may try to exercise the exclusive rights linked to the intellectual property right concerned or keep the information secret. Any unreasonable exercise of rights may stifle data sharing and thus innovation through big data, including in the transport sector. This is however mainly due to the inherent nature and purpose of intellectual property rights and trade secrets protection, which may at the same time provide an incentive for stakeholders to engage in data sharing for big data purposes.

A full analysis of the issues related to intellectual property rights is included in Deliverable D2.2, Section 3.7, pp. 135-177.

Table 14 Barriers and limitations: Intellectual property rights

Code	Description	Source(s)
LIM-LEG-IP-1	From a copyright perspective, even if the originality threshold of works is relatively low, most data used in the context of big data projects will not be considered original. It can thus not be assumed that such data will benefit from copyright protection, which leads to legal uncertainty.	New (inspired by LeMO D2.2)
LIM-LEG-IP-2	Copyright protection requires obtaining authorisation from the copyright holder of each individual data (as long as such data is indeed protected by copyright). In the context of big data projects, this may mean identifying authors of hundreds (if not hundreds of thousands) of works. In many cases, it might be difficult to identify or find the right holder and/or understand whether he has given his authorisation for the use of the work. In practice, this means that time-consuming analyses need to be performed before the data gathered can be used.	New (inspired by LeMO D2.2)
LIM-LEG-IP-3	From a copyright perspective, taking into consideration that the EU legal framework for copyright does not provide for a registration system, the eligibility for protection (and its scope) can only be confirmed *a posteriori* by a court, leading to a lack of legal certainty in the meantime.	New (inspired by LeMO D2.2)
LIM-LEG-IP-4	The "exclusive" character of copyright protection constitutes a hindrance, since in principle it does not allow acquiring copyright in the same data "in parallel". The copyright protection foresees for the work to have one author or several co-authors but in principle excludes the possibility that different entities acquire the same right independently under a different title (e.g. if the data were collected independently or on the basis of different sources). The latter may however often be the case in a big data context, in particular where parties will be independently collecting the same or similar data, leading to the creation of convergent datasets.	New (inspired by LeMO D2.2)
LIM-LEG-IP-5	In some Member States, there is no possibility to validly assign moral rights, which therefore requires taking additional measures to guarantee that the acquirer of the	New (inspired by LeMO D2.2)

	economic rights is free to use and modify data protected by copyright, to the extent necessary for big data projects.	
LIM-LEG-IP-6	It is unclear how data in contexts such as IoT, AI, algorithm- and sensor-generated data, and big data are protected by a database right (e.g. whether the current definition of a database embraces such technologies, who makes and who owns the database, and what 'substantial' investments are). A priori, the Database Directive does not apply to databases generated by means of machines, sensors, and other new technologies (such as IoT or AI).	New
LIM-LEG-IP-7	The *sui generis* right provided under the database protection legislation allows for the protection of the investment made in obtaining, verifying or presenting the contents of the database. The characteristics of such protection have however reached their limits in the current data- and technology-rich landscape as they do not sufficiently account for the following developments: (i) extended and intensive use of the internet and increasing economic value of data; (ii) distinction between efforts devoted to different database-related tasks (e.g. data creation, collection, arrangement, update, maintenance, publication); (iii) aggregation of data and big data; and (iv) automatic data generation; and advanced computational methods for analysis, information and decision making.	New
LIM-LEG-IP-8	Copyright protection of the structure of a database requires establishing that such a database is "original" by reason of the selection and/or arrangement of its contents. This requires an in-depth analysis of the internal structure and architecture. Hence, such protection is hard to establish with certainty.	New
LIM-LEG-IP-9	The level of and threshold for intellectual property protection diverge across the EU, which may hinder the possibility to manage pan-European big data projects, since it implies the necessity to examine multiple national legislations in order to have clearance on the possibility to use data, or secure the investment made in a database containing data originating from different territories.	New (inspired by LeMO D2.2)
LIM-LEG-IP-10	The recent EU copyright reform introduced exceptions (covering both copyright and database right), including one	New

	related to the use of data for the purposes of text and data mining on condition that the use of works and other subject matter has not been expressly reserved by their right-holders. It remains however unclear how this last condition is to be interpreted and how it will apply in practice.	
LIM-LEG-IP-11	Trade secret protection allows for protection of individual pieces of information regardless of their originality but requires the data to remain secret, which would hinder data sharing.	New (inspired by LeMO D2.2)
LIM-LEG-IP-12	It may be difficult to demonstrate that an individual piece of data benefits from trade secret protection, in the sense that it has commercial value because it is secret (many data will be considered valuable only if they are part of a bigger dataset).	New (inspired by LeMO D2.2)
LIM-LEG-IP-13	Many different actors in the big data analytics lifecycle may try to claim intellectual property rights in (parts) of the datasets intended to be used and may therefore try to exercise the exclusive rights linked to the intellectual property right concerned. Any unreasonable exercise of rights may stifle data sharing and thus innovation through big data, including in the transport sector.	LeMO D2.2
LIM-LEG-IP-14	Even in circumstances where no intellectual property rights subsist in individual data or in datasets, an assertion of (non-existent) IPR claims by economically stronger actors in the big data analytics lifecycle and/or in the field of transport against those with fewer resources may in practice have effects equivalent to the exercise of strong intellectual property rights. Such behaviour would thus constitute a barrier to the uptake of big data in the transport sector.	LeMO D2.2

4.2.7 Open data

The 'big data' required to feed big data analytics tools typically emanates from a variety of sources. One such source is the public sector, which has been opening up certain of its datasets to the public. The EU institutions have taken both legislative and non-legislative measures to encourage the uptake of open data, most notably through Directive 2003/98/EC on the re-use of public sector information (the PSI Directive), which attempts to remove barriers to the re-use of public sector information throughout the EU. Still, open data regimes also encounter a number of challenges – on a technical, economic and legal level – that cannot be ignored. The proposal for a recast of the PSI Directive aims to address some of these concerns. A major change concerns the expansion of the Directive's scope to include public undertakings. While information sharing has not been made mandatory for public undertakings (yet), the new regime constitutes a significant development for the transport sector, where services are often provided by public undertakings.

A full analysis of the issues related to open data is included in Deliverable D2.2, Section 3.8, pp. 178-196.

Table 15 Barriers and limitations: Open data

Code	Description	Source(s)
LIM-LEG-OD-1	Although standard open data licences are encouraged by the PSI Directive, it is shown that in practice, licences are still widely diverging in different Member States, often relying on US-driven licensing schemes. As a consequence, any company that wishes to reuse PSI, notably in a big data context, from different Member States, with the aim of developing a product is obliged to take into account a multitude of licences.	LeMO D2.2
LIM-LEG-OD-2	Public bodies are faced with the difficult task of reconciling their obligations under the PSI Directive and requirements under the data protection legislation. Data protection legislation presents a challenge to the opening up of public sector information, either because it risks preventing a large part of PSI datasets from being disclosed altogether or because it creates compliance issues when public sector bodies do decide to disclose PSI containing personal data.	LeMO D2.2
LIM-LEG-OD-3	Legal uncertainty exists about the precise relationship between the PSI Directive and the Database Directive. The PSI Directive states that it is without prejudice to that Directive and excludes from its scope all documents "for which third parties hold intellectual property rights". This has been frequently relied upon by public bodies to exclude the applicability of the PSI Directive to their information and thus allowing them to circumvent the rules of the PSI Directive even where the data is perhaps not actually covered by any intellectual property right.	New (inspired by LeMO D2.2)
LIM-LEG-OD-4	The example of essential services and critical infrastructures (e.g. railway infrastructures) shows that there are limits to	LeMO D2.2

	the desirability of open data policies, which should be taken into account by the EU legislator in current and future reviews of the PSI Directive.	
LIM-LEG-OD-5	Openness of (types of) data may differ between Member States, limiting in certain circumstances the scope of big data applications, including from a territorial point of view.	LeMO Case Study 4

4.2.8 Data sharing obligations

There exist legal instruments that impose specific data sharing obligations on private undertakings and therefore affect a company's control of, access to, or use of data. Such legislations are usually sector-focused and provide for an array of rights and obligations in relation to specific types of data in particular circumstances. Some of those pieces of legislation imposing data sharing obligations are particularly relevant to the transport sector, where for instance data sharing obligations are increasingly adopted in the context of Intelligent Transport Systems. The EU should however carefully consider whether the imposition of such general data sharing obligations is in each case equally necessary.

A full analysis of the issues related to data sharing obligations is included in Deliverable D2.2, Section 3.9, pp. 197-210.

Table 16 Barriers and limitations: Data sharing obligations

Code	Description	Source(s)
LIM-LEG-DS-1	Data sharing obligations are increasingly adopted in the context of Intelligent Transport Systems (ITS extends to all modes of transport, including rail transport, maritime transport and transport using inland waterway, and air transport). The adoption of data sharing obligations through legislations at EU and national levels is not necessarily the adequate way to force data sharing which in certain cases could be imposed through alternative means, such as public tenders or other incentivising schemes.	New (inspired by LeMO D2.2)

4.2.9 Data ownership

If the numerous stakeholders in the (big) data analytics lifecycle cannot rely on any of the other exclusive rights (such as in particular intellectual property rights), they increasingly try to claim "ownership" in (parts of) the datasets used in the analytics. No specific ownership right subsists in data, and the existing data-related rights do not respond sufficiently or adequately to the needs of the actors in the data value cycle. Up until today, the only imaginable solution is capturing the possible relationships between the various actors in contractual arrangements. Nevertheless, filling the data ownership gap with contractual arrangements is far from ideal from a legal perspective.

A full analysis of the issues related to data ownership is included in Deliverable D2.2, Section 3.10, pp. 211-223.

Table 17 Barriers and limitations: Data ownership

Code	Description	Source(s)
LIM-LEG-DO-1	While there is no specific ownership right in data, the existing data-related rights do not respond sufficiently or adequately to the needs of the actors in the data value cycle. Accordingly, the actors currently fill the data ownership gap through contractual arrangements.	New (inspired by LeMO D2.2) LeMO Case Study 4 LeMO Case Study 7
LIM-LEG-DO-2	Despite the absence of specific ownership right in data, multiple actors involved in the big data (analytics) value chain may try to claim ownership in (parts of) a dataset, which may hinder the production of, access to, linking and re-use of big data, including in the transport sector.	LeMO D2.2 LeMO Case Study 1 LeMO Case Study 4

4.2.10 Data sharing agreements

Currently, in practice, data sharing agreements are relied on to govern the access to and/or exchange of data between stakeholders in a big data analytics lifecycle. It is unclear, however, whether such practice enables covering all possible situations with the necessary and satisfactory legal certainty. Indeed, data sharing agreements entail numerous limitations in the absence of a comprehensive legal framework regulating numerous rights (e.g. ownership, access or exploitation rights) attached to data, the way in which such rights can be exercised, and by whom.

A full analysis of the issues related to data sharing agreements is included in Deliverable D2.2, Section 3.11, pp. 224-235.

Table 18 Barriers and limitations: Data sharing agreements

Code	Description	Source(s)

LIM-LEG-DSA-1	The complexity of data flows as well as the multitude of actors, data sources, algorithms, analyses, and activities that can be performed on data in a big data context requires the conclusion of a myriad of intricate data sharing agreements.	LeMO D2.2
LIM-LEG-DSA-2	Contractual agreements 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.	LeMO D2.2
LIM-LEG-DSA-3	It proves extremely difficult to clearly define the concepts of "data" and "data ownership" in data-sharing agreements as no legal definitions of these concepts exist.	LeMO D2.2
LIM-LEG-DSA-4	The actual terms of the data-sharing agreement may be difficult to negotiate depending on the parties' respective negotiating powers and may lead to having restrictive terms that in essence overly limit what the recipient may do with the data.	LeMO D2.2 LeMO Case Study 1 LeMO Case Study 4 LeMO Case Study 5 LeMO Case Study 6
LIM-LEG-DSA-5	Governing the big data analytics cycle through multiple data sharing agreements requires integrating complex back-to-back warranty clauses in respect of the upstream data sources as well as the downstream uses of data. In the absence of such clauses, the use of data may be prohibited or restricted, and the whole big data analytics chain may be affected.	LeMO D2.2

4.2.11 Liability

The current status of contractual liability rules, which may differ across the EU, is likely to limit the uptake of new technologies, including big data in the transport sector. The EU institutions have been engaged in ongoing work regarding extra-contractual and statutory liability in the context of disruptive technologies. On such basis, it will be possible to determine whether regulatory intervention is required.

A full analysis of the issues related to liability is included in Deliverable D2.2, Section 3.12, pp. 236-249.

Table 19 Barriers and limitations: Liability

Code	Description	Source(s)
LIM-LEG-LI-1	There may be a myriad of liability issues arising from different situations in a big data context (e.g. from inaccurate data), whereby the parties in contractual arrangements will try to limit, disclaim, exonerate their liability or transfer it onto the other party.	LeMO Case Study 2 LeMO Case Study 5
LIM-LEG-LI-2	In general, 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.	LeMO D2.2

4.2.12 Competition

As such, big data aggregation in the transport sector can give rise to a variety of competition law issues that suggest that certain aspects of competition law may not be fit for purpose. Abuse of dominance, merger control and anticompetitive behaviour have all seen challenges in the face of big data, AI and digitisation. The recent public consultation on shaping competition policy in the age of digitisation has yielded some interesting insights on how to mould competition law to address these topical issues.

A full analysis of the issues related to liability is included in Deliverable D2.2, Section 3.13, pp. 250-276.

Table 20 Barriers and limitations: Competition

Code	Description	Source(s)
LIM-LEG-COM-1	Recent changes to national merger control rules (Germany, Austria) to take account of "deal size thresholds" rather than turnover figures could result in more data-rich mergers requiring prior merger clearance. A similar approach at EU level is under consideration.	LeMO D2.2
LIM-LEG-COM-2	Competition compliance programmes may need to be examined to verify whether price-fixing could result from the coordinating effect of algorithms and associated risks.	LeMO D2.2
LIM-LEG-COM-3	Refusals to share data may in certain cases raise abuse of dominant position concerns, especially when considering leading technology companies (e.g. GAFAM).	LeMO Case Study 6

5 Ethical and social barriers & limitations

Many social and ethical aspects such as transparency, profiling and tracking, re-use, sharing, open data and open access require consideration in big data practices in the transport sector. In addition, big data practices concern data from people, therefore the human element is associated with individual social and moral codes. Further, these aspects require recognition so that public and private transport organisations can incorporate fundamental ethical and social values into big data practices and policies. To support the European big data transport industry in socially and ethically responsible practices in the transport field, this section summarises ethical and social barriers/limitations, which have been extracted from LeMO and other initiatives (i.e., NOESIS and TT projects).

First of all, social and ethical aspects considered under the LeMO project are briefly described as follows[4]:

- **Trust**: As one of the main dimensions of big data describing consistency and trustworthiness, "veracity" is related to trust. It could be classified into two aspects: 'trust for big data' related to trust and reputation systems and 'trust in big data' measuring the trustworthiness and accuracy of big data to create high values.
- **Surveillance**: In common perception, it is used in crime prevention or criminal investigation. However, surveillance in the transport sector and big data tends to be a close observation of all human behaviour in general, irrespective of their criminal tendencies.
- **Privacy**: Despite the difficulty to define this concept, since it covers many different dimensions, it could be presented as "a state in which one is not observed or disturbed by other people".
- **Free will**: Free will is usually considered as distinct from other concepts such as 'autonomy' and 'authenticity'. Also, it can be defined as a kind of power or ability to make decisions of the sort for which one can be morally responsible.
- **Data ownership**: Ownership of data (whether it is personal or non-personal) has been heavily debated throughout the EU and in other parts of the world and means control capability over the collection, processing and sharing of personal data according to GDPR in May 2018.
- **Social discrimination**: It could be grouped into unintended and intended data biases. The former results from wrong statistical treatment or poor data quality at any step of big data analytics pipeline. The latter occurs due to intentional filtering of data and knowledge on customer or user behaviour and access.
- **Environment**: There are trade-off or rebound effects, which are related to energy efficiency and emission gas in the transport sector, limiting the effect of big data exploitation or creating unintended consequences, in terms of environmental aspects.

More details about the above ethical and social aspects can be found in Deliverables 2.3 and 2.4 of the LeMO project.

---

[4] Deliverable 2.3 of LeMO project

In the evaluation of NOESIS, the data ownership and user rights are represented in only 7 and 3 among 25 cases[5]. The data ownership could be a problem when the data is related to personal information as figured out by the four case studies (i.e., CS2, CS3, CS4, CS5 and CS7) in LeMO project. Whereas, the social discrimination can be considered as a side-effect of using high technologies as mentioned in two case studies CS1 and CS7.

Table 21 Barriers and limitations: Data ownership

Code	Description	Source(s)
LIM-ES-DO-1	There are conflicts about data ownership between commercial activities. That is, multiple actors involved in big data analytics may try to claim ownership of the data concerned, which may lead to a gridlock.	LeMO Case study 2 and 7 NOESIS (11 use cases)
LIM-ES-DO-2	There are threats for user rights of personal data ownership. As the definition of 'personal data' continues to evolve, information that is qualified as non-personal data today may be classified as personal data in the future. 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.	LeMO Case study 3, 5 NOESIS
LIM-ES-DO-3	End-users may be reticent to provide their personal data for big data analytics in transport as this would entail forsaking "ownership".	LeMO D2.3
LIM-ES-DO-4	Difficulty in establishing ownership of different data components within a set of data of various types and coming from various sources.	LeMO D2.3

Table 22 Barriers and limitations: Social discrimination

Code	Description	Source(s)

[5] See the frequency charts at the page number 57 of Deliverable 2.3 of NOESIS project.

LIM-ES-SD-1	There may be discriminations of levels for offering and utilising information between users and companies ("Digital Divide").	LeMO Case study 1
LIM-ES-SD-2	Personalised services could exclude certain social groups or lead to discriminatory treatment.	LeMO Case study 7
LIM-ES-SD-3	Big data analytics are vulnerable 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 algorithms complexity.	LeMO D2.3

For the **trust and environmental aspects**, there are more optimistic views about the use of big data in the transport sector to ensure the trust of data and services and to save the environment, than when it comes to the other aspects such as privacy, surveillance and free will. For example, cooperation for aggregating traffic data and using uniform standard could increase data trust. Also using big data analytics might have a positive impact on the greenhouse emissions, noise reduction, resource saving and reduction of animal run-over in the transport sector in Europe by operating with a minimal environmental impact of almost all modes. Indeed, the transport sector using big data might help to save the environment, as revealed in use case evaluation by NOESIS. In the investigation, the examples having positive effects varied as decarbonisation, pollution reduction, noise and vibrations reduction, and environmental monitoring. Among these effects, pollution reduction was the most frequently appearing in NOESIS use cases[6], and five systems (i.e., smart highways[7], smart connected vehicles[8], proactive rail infrastructures[9], smart airport turnaround[10] and integrated urban mobility[11]) out of seven have shown the effects[12]. However, we can also focus on negative effects revealed by the NOESIS and TT projects, such as reduction of space and resource by building transport infrastructure. Several Key Performance Indicators (KPIs) for the environmental aspect have been introduced and summarised by TT project[13].

Table 23 Barriers and limitations: Trust

Code	Description	Source(s)

[6] See the frequency chart at the page number 49 of Deliverable 2.3 of NOESIS project.
[7] See the Deliverable 4.1 and 4.3 of TT project.
[8] See the Deliverable 5.1 and 5.3 of TT project.
[9] See the Deliverable 6.1 and 6.3 of TT project.
[10] See the Deliverable 8.1 and 8.3 of TT project.
[11] See the Deliverable 10.1 and 10.3 of TT project.
[12] See the environmental quality category at the page number 38 of Deliverable 2.2 of TT project.
[13] See the Deliverable 2.2 of TT project.

LIM-ES-TR-1	Low quality of data due to error and heterogeneous values can harm the trust of users for using big data and its technology. It also happens when there are redundant data between various types of data.	LeMO Case study 2, 3, 5 and 6 NOESIS and TT (Many cases)
LIM-ES-TR-2	Old and un-updated data reduce the quality of data and the trust for data.	LeMO Case study 2 NOESIS
LIM-ES-TR-3	Trust related to big data can be easily polluted by various kinds of people who act against a moral requirement.	LeMO D2.3

Table 24 Barriers and limitations: Environment

Code	Description	Source(s)
LIM-ES-ENV-1	Even though there are positive effects such as reduction of pollution emission, animal run-over, noise, vibration and decarbonisation by using big data technology, increase of ICT infrastructure could create rebound effects.	LeMO Case study 5 and 6

Surveillance and privacy aspects have also been presented in the evaluation of NOESIS[14]. In this regard, negative concerns seem to be high due to being considered as personal information restricted by GDPR regulations. Many negative issues occur by misusing data such as monitoring passengers and drivers, tracking people.

Table 25 Barriers and limitations: Surveillance

Code	Description	Source(s)
LIM-ES-SUR-1	People are concerned about surveillance due to collection of data, which is likely to include individual information. They believe this would threaten their privacy.	LeMO Case study 2, 3 and 6 NOESIS (4 use cases)

[14] See the frequency chart at the page number 57 of Deliverable 2.3 of NOESIS project.

LIM-ES-SUR-2	An individual could be identified by aggregated big data analysis. And it is one of the biggest concerns on the utilisation of big data in the transportation sector also.	LeMO Case study 3, 6 NOESIS
LIM-ES-SUR-3	There are serious ethical issues posed by the emerging regimes of population-level monitoring, whereas recent privacy-protection initiatives fall short of addressing the challenge to democracy posed by big data surveillance.	LeMO D2.3
LIM-ES-SUR-4	If predictive analytics succeed in altering behaviours, surveillance would no longer remain a passive technique but would become a dynamic tool to shape behaviours (impacting for instance free will).	LeMO D2.3
LIM-ES-SUR-5	Various commentators consider that the privacy risks related to big data analytics are low, pointing out the large amount of data processed by analytics and the de-identified nature of most of this data. This conclusion is likely to be wrong in practice, including from a legal perspective. This is notably due to the fact that anonymity by de-identification is a difficult goal to achieve, as demonstrated by different studies.	LeMO D2.3

Table 26 Barriers and limitations: Privacy

Code	Description	Source(s)
LIM-ES-PRI-1	Privacy is a big issue as data from transportation is quite close to how people live, and it might include personal information.	LeMO Case study 2, 3 and 6 NOESIS
LIM-ES-PRI-2	With the increase of investments and expectation to apply IT in transport systems, vehicles become increasingly automated. This further complicates the privacy issues, since vehicle data gets mixed with personal data. Even if only authorised actors are allowed to access and use this data, still it remains to be a major concern.	LeMO Case study 3, 5 NOESIS
LIM-ES-PRI-3	As needs for privacy vary between individuals or between situations (e.g., depending on the benefits the individual	All case studies

	gets in return), it will be difficult for companies and developers to adopt a one-size-fits-all approach.	

Free will aspect has various kind of issues on the utilisation of big data in the transport sector. Positive issues might be controlling the traffic flow to save fuel and be more punctual or helping to shape the travellers' choices. Whereas, people's opinions could be ignored when the area uses and analyses big data. This aspect has shown more neutral position in the LeMO case studies (i.e., railway transport, open data and traffic management fields), than the other aspects.

Table 27 Barriers and limitations: Free will

Code	Description	Source(s)
LIM-ES-FW-1	Free will could be threatened by low quality of transportation information, for example, an improper decision about traffic restrictions based on analysis of traffic data will waste time resources and return with large amounts of claims.	LeMO Case study 5
LIM-ES-FW-1	The rapidly increasing size and scope of information of big data technologies could lead to unfair use, in terms of the free will and autonomy of humans.	LeMO D2.3

Table 28 Barriers and limitations: Others (i.e., transparency, consent and control)

Code	Description	Source(s)
LIM-ES-OT-1	**Transparency:** The lack of transparency of personal data processing activities in a big data context negatively affects data subjects' trust in such activities and the related technology. Data subjects may be reluctant to use big data applications in the transport sector.	LeMO D2.3
LIM-ES-OT-2	**Consent:** The misconceptions regarding data protection concepts, such as consent, cause confusion both among data subjects and organisations. This general trend will also affect the use of big data in the transport domain.	LeMO D2.3
LIM-ES-OT-3	**Control:** There is an asymmetry of control of personal data between data subjects and the organisations processing the data. Data subjects may fear losing control over their digital identity by engaging in big data analytics.	LeMO D2.3

6 Economic and political barriers & limitations

In Lemo, the framework presented in the figure below was used to understand the interactions between the key transport stakeholders from the perspective of the big data economy.

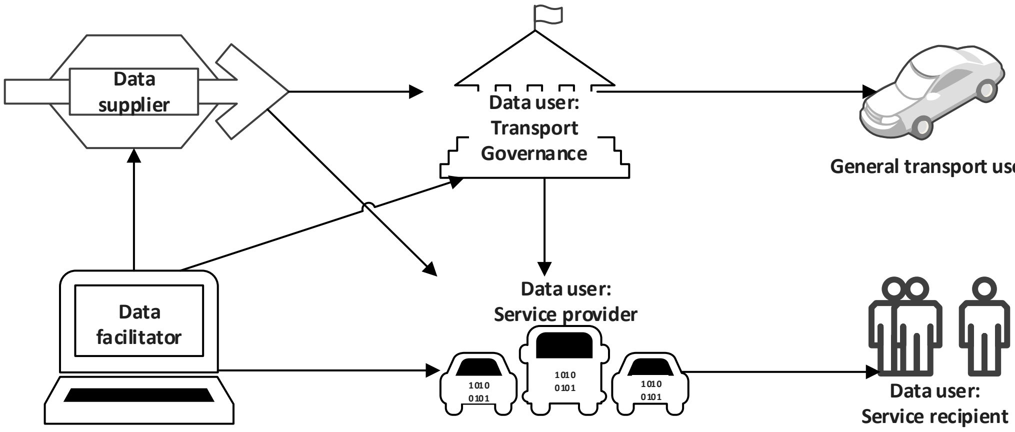


Figure 3 Framework for understanding interactions between stakeholders

Previous work in LeMO has established that there are three main data user groups:

1. Service providers, ranging from vehicle drivers to network operators or even advisory or standalone information services.
2. Service users, who benefit from the service provider's use of big data technology.
3. Transport governance, primarily public authorities who regulate and make high impact decisions on how transport services operate and how traffic infrastructure is built.

The decisions made by transport governance affect primarily transport service providers and general transport users. However, a new political role might also need to be defined for governing the service providers reliant on big data.

Supporting the use of big data in the transport sector are the roles of data supplier and data facilitator. Data suppliers are primarily from the transport sector themselves, but may also come from other industries, such as supply chain actors, real estate developers, vehicle manufacturers, mobile phone app service provider, banking sector, or even the telecommunications industry. Naturally, public authorities also collect a lot of relevant data, and may "share" it through restricted or open data portals.

Data facilitators, on the other hand, generally originate from the ICT sector, and are integral in providing the tools for collecting, storing, cleaning, processing visualizing and analysing big data. In some cases, conventional data management companies can be used, but the technical requirements of big data may necessitate unique big data technology. Furthermore, some niche companies have developed big data analysis techniques based on machine learning and artificial intelligence principles. In many cases, tool development has also originated with the data users who have a better understanding of the context and the mindset of the needs of the company.

Learning from work done in WP2 and WP3 of LEMO and the results thus far in the NOESIS and TT projects, we have identified barriers to the use of big data in the transport sector. Underlying these barriers are 5 mechanisms or types of influences on the big data in transport activity. These mechanisms are briefly titled:

- Inaccurate results
- Restricted access to data
- Lacking institutional capacity
- Underdeveloped big data industry
- Uncertain political risk

6.1.1 Mechanism 1: Inaccurate results

The primary motivation for any application of big data technology is explicitly the demand for better informational and analytical content. This affects the competence of data users to carry out their activities, whether as service provider, service user or in the role of governance. Unless factors that cause inaccurate results can be removed or its negative effects limited, it will be difficult to justify the adoption of big data technology in the sector. Businesses will have to judge the benefits also on the basis of opportunity costs, i.e. to evaluate other potential technological options or to simply not adopt any new system.

Different types of issues were identified that have a similar general effect that is inaccurate results. While there are many more factors that contribute to inaccurate results, the two highlighted here are considered core issues:

EP-IR-1: Bias from data sample and algorithms.

EP-IR-2: Bad quality of data.

Bias from data sample and algorithms.

Data scientists have highlighted the fact that algorithms are not "value-free", but are heavily dependent on the choices that the developers themselves have made. The two main factors here are the bias inherent in the big data itself and the algorithms that read, store and analyse that data. Both primarily reflect the challenge of selecting the right technological solution, ranging from data collection methods and technology to the intelligence underlying the analysis. Selection of technological solutions in a business setting does not only depend on the desired capabilities of the solution, but also on the cost of acquiring the solution. The total cost of ownership[15] paradigm used in decision making processes for acquiring assets provides a good

[15] Ellram, L.M., 1995. Total cost of ownership. Int. J. Phys. Distrib. Logist. Manag. 25, 4–23. https://doi.org/10.1108/09600039510099928

framework to understand the various cost components, ranging from the investigation of options, supplier selection, acquisition, up to the use and disposal of systems (and data).

The main outcome of the bias is the inaccurate understanding of the market or more generally, the subject of analysis. This will lead to poor design and execution of the business model. In some cases, the business may provide a socially discriminatory service, which is also damaging to its reputation. The direct recipients of the negative impacts are general transport users, externality recipients and the service recipients. As a result of providing the poor service, the service provider and transport governance are impacted financially, and perhaps to its reputation. Another aspect of the issue is that bias may be difficult to detect by the company, until it is significant and costly decisions have already been made.

However, it may nevertheless be rational to stick to these biased methods, if the cost of switching to a *better system* is not viable either in the short or long term. In this case, the effects of the inaccurate results could be managed in other ways, for instance, by supplementing the results with other means of analysis or expert opinion.

Further, a strategic question is whether the company should persist in using and refining the big data methods despite the known poor results. This is a risk, but one which businesses nevertheless have to undertake to be in the "best" position for exploiting the technology (and data) once the bias has been dealt with.

Bad quality of data.

Bad quality of data stems from internal technological or data governance issues, but also from decisions made by data suppliers. Inaccurate and unreliable technology used to collect, store and transmit data can cause systematic errors in the data quality. If data governance procedures are not missing or poorly executed, bad quality data can remain undetected until it becomes very costly to correct. Bad quality data can also be passed on from a data supplier to a data user.

Similar to biased results, bad quality data can lead to poor decisions made. However, not necessarily leading to discriminatory services. Bad quality data should presumably be easier to detect than bias. In any case, it is difficult to justify using bad data, in the absence of good data, because the results can be more unreliable.

Table 29 Barriers and limitations: Inaccurate results

Code	Description	Source(s)
LIM-EP-IR-1	Bias from data sample and algorithms are difficult to detect. Besides yielding poor analytical outcomes, it may also lead to unintentional socially discriminatory services. Since it is difficult to detect, data users may decide not to change their operations, as long as the business does not suffer from financial or reputation loss.	LeMO Case study 7 LeMO D1.1, D2.1, D2.2, D2.3 NOESIS D2.3

LIM-EP-IR-2	Bad quality and inappropriate data lead to poor analytical results, negatively impacting the business model. Consequences depend on the use of the analysis, ranging from inaccurate market understanding, poor business execution to bad products for customers.	Case study 2, 3, 5 LeMO D1.1, D2.2, D2.3

6.1.2 Mechanism 2: Restricted access to data

Supply of data fall under three main categories of ownership from the perspective of the organization:

(1) Self-owned: where the supply of the data is from within the firm.
(2) External commercial ownership: where the rights to the data are held by another organization and must be transferred to the organization to be used.
(3) Open data: where the rights of the data is has an open data license.

For the data to be useful to the organization, restrictions to the acquisition of (sufficiently) high quality data must be addressed. Failure to do so leads to the collapse of the business model, high costs of data acquisition, and poor value extraction of the data. If the organization cannot ensure the flow of data from external sources (i.e. supply (2) and (3)), whether via legal or technological means, it may consider to develop and control its own source of data. Technological issues are addressed in Section 3.

The three issues highlighted here are:

EP-RD-1: Perception of safety and privacy.

EP-RD-2: Emergence of data silos.

EP-RD-3: Monopoly of data sources.

Perception of safety and privacy.

There is often a trade-off with perceived usefulness of the service, the perceived concerns regarding safety, security and privacy, and the costs of using the service. The perception of safety, security and privacy has been studied together with the rise of digitalization in various industry sectors, such as banking and healthcare, using the Technology Acceptance Model framework. The two main concerns are on the use of data for unintended purposes, especially in the case of data breaches or mismanagement, and the development of personal profiles, which are perceived to be “intrusive” into the personal lives of the consumers.

The first primarily affects the question of whether the person and their belongings are safe and secure in the case that my data ends up in the hands of criminals. This issue is important to both private persons and organizations. Organizations that outsource their data management might also have the added issue of being nevertheless responsible to their customers for securing their data. The second primarily affects the private person that might be concerned that a particular organization or the government has insight into the behaviour, attitudes and personal life of themselves or their personal relations.

The types of data generally associated with this issue is the collection of GPS data, public transport travel information (e.g. from fare cards), e-commerce transactions, websites visited, and financial transactions. The advertising business units of Alphabet and Facebook entice users with free services, in exchange for access to data to create targeted ads. However, as a platform, it can also rely on other streams of information or rely on more generic user profiles in the ads. Currently, big data-aided transport services use the data collected to improve the services provided, not for side-income. Hence, it strictly relies on good data, as discussed in Section 3.2.1.

A poor perception of safety, security and privacy can lead the user to stop to use the service or to use the service but provide misleading information. The issue of trust (refer to Section 5 “Trust”) becomes vital to the survival of the business model.

On the other hand, it has also been noted that the use of big data may supplement other more intrusive data sources that yield the equivalent results needed for planning. Hence, the type of information needed yields a group profile, rather than a personal profile. Both are significant improvements from the current data collection methods relying on individual surveys.

Emergence of data silos.

Data silos are generally an ex post problem, where a unit of an organization realizes that it needs certain data, which is held by another unit of the same organization. However, past technological, legal or organizational decisions restrict data from being transferred to requester. The ex-ante intention may not have been to restrict sharing of data with others within the organization, but due to historical use of incompatible technology, observing the policy or terms set out on the collection and use of the data, and perhaps simply the lack of awareness of available data in the silo.

If it is allowed, data silos can be “dismantled” step by step, depending also on the type of barrier. Dismantling the silos might require oversight and cooperation across departments, which might be organizationally difficult. Further, it may also be the question of assigning the costs internally within the organization, which will reflect on the performance of the individual units.

Monopoly of data sources.

Census activity, infrastructure operation, licensing and enforcement have historically yielded population-level data sets on vehicle ownership, vehicle use, and travel activity. However, the rise of digital-first companies have in some cases supplanted the privileged position of public authorities in owning data on transport sector. For instance, the data owned by Waze (or Alphabet, the parent company) or Uber often exceeds the richness of data available to transport authorities. The data they have collected is extremely valuable and forms a part of their business model as data suppliers. Further, as many services become “natural” monopoly or oligopoly, in the sense, that only a few services dominate the market share, the data owned in that space becomes more difficult to source or collect again. For instance, the information from a car route planner collected by Waze (or Google Maps, Apple Maps, etc.) can presumably only be collected at that precision level (i.e. 1 Hz GPS coordinates) by another route planner, i.e. not from traffic sensors or cameras. While difficult, alternative data sources, such as telematics services of a vehicle, could also be used for similar purposes. After all, a monopoly of a service does not imply a monopoly of that data or information.

Nevertheless, it is certain that data, often because it is rich, contains sensitive information and is difficult to collect, naturally becomes valuable and difficult to access. Data collectors have to be entrepreneurial to create alternative data collection methods usually encapsulated in a business model that serves the desired data objects. Business models that fail to create value in serving the data objects will not succeed in data collection as well.

Organizations that insist on using data held by another must make an equivalent exchange (whether financially or services in-kind). This is not unusual in any commercial activity. However, public agencies may become dependent or be unduly influenced by commercial companies for their data.

Table 30 Barriers and limitations: Restricted access to data

Code	Description	Source(s)
LIM-EP-RAD-1	Data subjects are right to be concerned about the breach of safety, security and privacy, if their personal data is collected. Actions by the organization or perhaps in the news that affects their perception of safety, security and privacy may affect the willingness of the data subject to "submit" their data. On the other hand, big data methods may be applied to circumvent the lack of some kinds of data.	LeMO Case study 3, 4, 7 NOESIS D4.3
LIM-EP-RAD-2	Data silos restrict the combination and re-use of data. Several reasons, such as historical choice of incompatible technology, observing the policy or terms set out on the collection and use of the data, and perhaps simply the lack of awareness of available data in the silo. Silos are important to maintain to preserve privacy and other intentional data recombination restrictions.	Case study 2, 4 LeMO D1.1, D2.1
LIM-EP-RAD-3	Data, often because it is rich, contains sensitive information and is difficult to collect, naturally becomes valuable and difficult to access. New digital-first service providers have better access to quality data. Business models that fail to create value in serving the data objects will not succeed in data collection as well. Providing value is key for a thriving business and should be guarded. Alternative data capture methods can be pursued.	NOESIS D2.2

6.1.3 Mechanism 3: Organizations lack capacity

While big data techniques are still developing, an important barrier is the institutional capacity of all big data actors (i.e. users, suppliers, and facilitators) that are lagging behind. Relevant aspects of institutional capacity are in (1) the knowledge and competence of data workers and the business developers; (2) internal organisational structure, policy and strategy; and (3) informal and formal institutions, policy and regulatory framework. The first two are focused on the business organization and "value configuration". The third focuses on both the formal and informal institutional arrangements that restrict or impose certain actions on the organization, failing which, the organization usually receives a penalty, ranging from intangible, such as the loss of reputation, to the tangible, monetary fines.

The five issues highlighted here are:

EP-OC-1: Dark data is an unknown and uncertain risk.

EP-OC-2: Failure to contextualize and interpret the results of analysis.

EP-OC-3: Costly change management.

EP-OC-4: Shortage of skilled labour.

EP-OC-5: Data governance is not implemented.

Dark data is an unknown and uncertain risk.

Dark data is the holding of data that poses risks to the company, without the knowledge of those risks. Companies that hold personal and sensitive information, whether of its own entity, of partners or other data objects, should be doubly aware of the risks that they pose. There are two major risks that we are concerned about here: (1) effects of data breach and (2) legal and regulatory risk. In type (1), a data breach of personal and sensitive information may lead to loss of reputation by the clients, loss of competitive edge, or also open the company to legal fines for failure to protect the sensitive information. In type (2), the holding of some data may not be permitted by law. If the company is found with the data, it may face severe penalties.

Failure to contextualize and interpret the results of analysis.

As organizations are "learning" to use big data technology and analytic methods, it may have faulty assumptions about the results and the conclusions one can make. This may have consequences similar to that of "inaccurate results", but it is a fundamentally different mechanism. Correctly performed analysis still requires a correct interpretation of the results in order to suggest actions. This is especially crucial when higher order and large impact decisions rely on big data analysis.

Skilled data scientists, i.e. those that not only perform analysis but also understand the models and processes behind it, are necessary to support the measured and safe growth of technology adoption. However, the industry is still new, and skilled data scientists are still lacking. Hence, many practitioners are self-taught without a strong scientific or statistical foundation. (This may be encouraged by the many online and free courses available.)

Costly change management.

Assuming a good business case, i.e. perhaps in the sense of a profitable Return-on-investment, it may still be very challenging to successfully bring the change necessary to the organization. Redesigning business processes and the corresponding retraining of staff are some of the necessary actions to properly integrate new technology in an organization. This may be costly, take a long time, and may face strong internal resistance. Human resources management becomes a key tool to bridge the necessary changes to the organization and the existing teams and their management.

Large organizations are generally more difficult to transition, while newer and smaller organizations may be more agile in this regard. On the other hand, large organizations have usually a larger budget to fund the change, including the investment of technology. Regardless, the adoption of big data technology by any organization should consider the potential of failing to transition.

Shortage of skilled labour.

Reliance on big data technology also requires data workers. This has been briefly mentioned in the previous two issues. Here, we expand a bit the scope of the issue. There are two main deficiencies in Europe that contribute to this shortage. The first limitation is the big data-based industry is still growing in Europe. Cluster efficiencies, such as found in Silicon Valley, is not present to the same level in Europe. Hence, there is limited growth in hiring opportunities for new data workers. Attractive job prospects are necessary to draw new graduates to get the relevant training. Students that have studied leave for other areas for better job prospects. The second limitation is the lack of integrated higher education training and studies. In turn, this affects location decisions of the companies. Hence, it forms a vicious cycle that restricts growth in the locality.

Nevertheless, job market issues can be easily solved with high salaries to attract talent. This contributes to the evaluation of the business case, whether the costs will exceed the potential revenue.

Data governance is not practised.

Data governance is not new. However, it is an essential practice and standard that becomes even more vital when organizations are to deal with big data. This issue is related to the issue of change management, where organizations need to spend resources to institute proper data governance. Relevant questions here are whether there are “standard” data governance approaches or more concretely practices and solutions. Organizations may not currently be equipped to start and may not have the oversight to maintain the governance systems. If this fails to be prioritized, issues such as dark data and data silos may emerge.

Table 31 Barriers and limitations: Organization lack capacity

Code	Description	Source(s)
LIM-EP-OLC-1	Dark data is an unknown and uncertain risk to the organization holding personal and sensitive information. A data breach of personal and sensitive information may lead to loss of reputation by the clients, loss of competitive edge, or also open the company to legal fines for failure to protect the sensitive information. The holding of some data may not be permitted by law. If the company is found with the data, it may face severe penalties. However, for an organization, having data is preferable to otherwise, since value may be extracted from dark data.	LeMO D2.1
LIM-EP-OLC-2	Self-taught data workers can quickly jumpstart a business, which is useful for a certain organization scale. Organizations are not equipped to properly contextualize and interpret results of analysis, needing skilled data scientists, besides just data analysts. The situation leads to similar outcomes as "Inaccurate results".	New
LIM-EP-OLC-3	Organizational and business process transitioning to rely on big data Technology-based processes is costly. Redesigning business processes and the corresponding retraining of staff are some of the necessary actions to properly integrate new technology in an organization. This may be costly, take a long time, and may face strong internal resistance.	LeMO D1.1 NOESIS D4.3
LIM-EP-OLC-4	Trained data analysts are lacking. The lack of data workers freeze the movement towards a vibrant big data economy. The lack of industries and education limits the training of data workers. Trained data workers, who leave the EU for employment, get overseas exposure and training.	Case study 1, 5
LIM-EP-OLC-5	Data governance within an organization is necessary but difficult to implement. Organizations may not currently be equipped to start and may not have the oversight to maintain the governance systems. If this fails to be prioritized, issues such as dark data and data silos may emerge.	NOESIS D4.2

6.1.4 Mechanism 4: Underdeveloped big data industry

A significant barrier is simply the progress of the technology and the market, accessible to Europe. This refers primarily to the ICT sector as suppliers, but also to the demand for this niche technology affecting the transport sector. While it is often conceptualized as a chicken-or-egg problem, a natural development of an industry usually follows more of a slow spiral, with the market experimenting with different technology applications and configurations. Even though the market eventually settles on a few technologies that become more dominant, the experimentation still continues with different variations in technology, application and problems.

The three issues highlighted here are:

EP-IU-1: Technology is still maturing.

EP-IU-2: Cost of infrastructure is expensive.

EP-IU-3: EU-based industry still lagging.

Technology is still maturing.

The development of technology and analysis methods are still maturing. As a niche technology, it is still unclear where it will be heading in the transport domain. Experimentation is required to reduce the uncertainty surrounding the capabilities of the big data-based technology. But, experimentation is also expensive when considering opportunity cost. Would another conservative choice of technology or method yield a better result at a lower cost? In reality, some industries would be more willing to experiment than others. Also, in some cases, there is funding available to test new technology, such as research funding.

But other restrictions of experimentation are costly access to data and lacking institutional capacity, which has been discussed elsewhere.

Cost of infrastructure is expensive.

Cost of owning and operating big data infrastructure is expensive. As an emerging field, it might need some time for the industry to grow more competitive and to learn/improve, in order for costs to reduce. The costs of purchase and operating big data infrastructure are high, unless it can scale up appropriately. A solid business case is a prerequisite for the investment. However, a reasonable Return-on-Investment (ROI) is difficult to guarantee. Smaller organizations will tend to remain conservative. Larger organizations may have the additional capacity to invest.

The “servitization” of big data infrastructure (e.g. cloud services) is an important way for smaller organizations to experiment and operate without costly investment in owning equipment. However, it also requires careful calculation to ensure that costs do not increase without monitoring.

EU-based industry still lagging.

On the whole, the big data industry in the EU is still lagging behind other nations. This makes it difficult for transport organizations to adopt the technology and practices, as clusters have synergies that are self-supporting. While solutions and expertise from other parts of the world can still support the EU-based industry, but it will be difficult for sustained growth. The slow growth has an overall effect on the other issues highlighted here.

Table 32 Barriers and limitations: Big data industry underdeveloped

Code	Description	Source(s)
LIM-EP-IU-1	The development of big data Technology, analysis techniques and application fields are still maturing. Experimentation is required to reduce the uncertainty surrounding the capabilities of the big data-based technology. But, experimentation is also expensive when considering opportunity cost. Big data is evaluated among other technological solutions in the market. Thus, investment is spread to "other" solutions, just in case big data does not deliver for the particular industry/company.	LeMO Case study 7 LeMO D2.1, D2.2, D3.2 NOESIS D4.3
LIM-EP-IU-2	Cost of owning and operating big data infrastructure is expensive. The costs of purchase and operating big data infrastructure are high, unless it can scale up appropriately. A reasonable ROI is difficult to guarantee. High costs lead to servitization of the infrastructure as a business model, which has a lower entry cost.	LeMO D1.1, D1.3, D2.1, D2.3
LIM-EP-IU-3	The EU lags behind other economies in the development of big Data Technology. It is difficult for transport organizations to adopt the technology and practices, as clusters have synergies that are self-supporting.	LeMO D2.1

6.1.5 Mechanism 5: Uncertain political risk

While bad results in a business setting, may bring about momentary economic losses (e.g. for failing to understand and serve the market correctly), the use of big data in the governance of public services and infrastructure has a certain political risk. This is therefore unique to the integration of big data technology in the carrying out the responsibilities of public authorities. The effects of an error in development and execution of policy affects citizens, which might backfire politically (e.g. in the loss of political power or popularity) and societally (e.g. in the loss of confidence in the government).

The key issue here is

EP-PR-1: Technocratic and algorithmic governance.

Technocratic and algorithmic governance.

Transport planning and policy making (and any policy making in general) must be the responsibility of policy makers and democratic representatives. If these actions rely too much on the use of automated and "black box" decision making processes, the culpability of transport governance actors reduces. Or more correctly, they may claim to be less culpable, since the decisions are the outcome of the technology and perhaps the data scientists that set up the systems. This would be a poor outcome for the democratic process, which is usually a negotiation and consensus-creating process. The key risk is that without representation, vulnerable and marginal groups, may not have a voice for their concerns. This issue is associated with the issue of biased results.

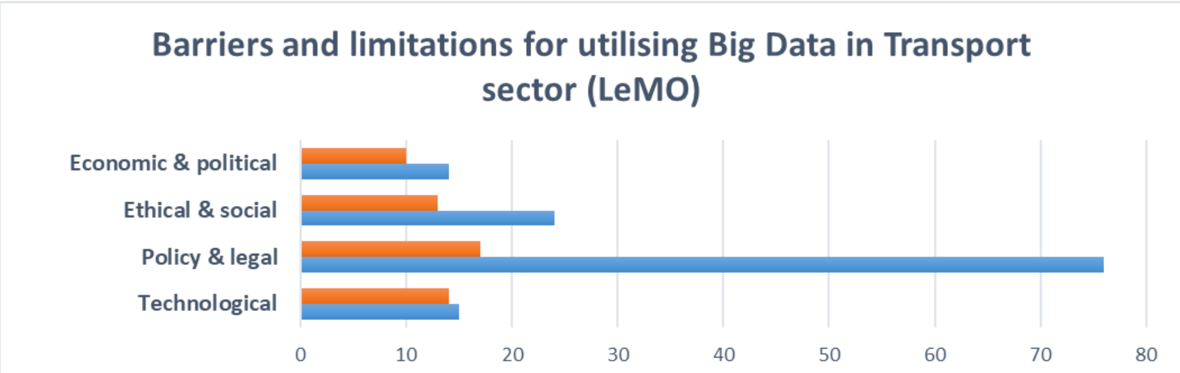
Table 33 Barriers and limitations: Uncertain political risk

Code	Description	Source(s)
LIM-EP-UPR-1	Extreme reliance on big data Algorithms may lead to technocratic and algorithmic governance of transport infrastructure. Without democratic representation, vulnerable and marginal groups, may not have the possibility to voice their concerns. Big data analytics and automated decision making may improve the efficiency and results of governance.	LeMO Case study 7 LeMO D2.1

7 Conclusion

In this deliverable, we uncovered different barriers to the utilisation of big data in the transport sector and their characterisation as technological; policy & legal; ethical & social; and economic & political. We also explored whether these barriers currently occur in real-world or not. In the course of this task, we considered the case studies of LeMO, the use-cases of NOESIS and the pilot systems of TT to gather a comprehensive set of barriers and limitations.

For each aspect/character (i.e. technological; policy & legal; ethical & social; and economic & political), the following figure summarises the numbers of the barriers and limitations investigated from previous work packages (i.e., WP1 and WP2) and discovered from our project cluster (i.e., LeMO, NOESIS and TT).



	Technological	Policy & legal	Ethical & social	Economic & political
# of barriers identified	14	17	13	10
# of barriers investigated	15	76	24	14

Figure 4 Summary of # of barriers and limitations for each aspect

From the samples studied under the three projects, it can be seen that technological, as well as economic & political aspects, are being considered and discussed more by the current industry, since the barriers falling under these aspects appear more often as compared to the other barriers (i.e. 93 and 73 percent of the times respectively)[16].

However, efforts to consider the other two aspects (i.e., ethical & social and policy & legal issues) are relatively low. This is because the barriers falling under these two aspects do not appear as often (only 54 and 22 percent of the times respectively).

It is also interesting to see that the potential barriers and limitations related to legal aspects are many and of various types (13 types of legal barriers have been discussed). It may be because more legal issues have evolved due to the recent introduction of regulations such as the GDPR, which have a substantial impact on the utilisation of big data and they are getting more attention nowadays.

[16] The percentage here reflects the number of barriers identified over the number of barriers investigated.

We figured out that there are links between the barriers falling under different aspects. This further complicates the impact of these barriers on the use of big data in transport. Furthermore, it was also observed that there are interventions that can diminish the negative impacts of the barriers or convert them into opportunities. Lastly, the barriers and opportunities also often mutually affect each other. These observations will be further discussed and explored in the forthcoming tasks in this work package.

References

[1] K. Hee *et al.*, 'D1.1 Understanding and Mapping big data in Transport Sector', 2018.

[2] T. Teoh *et al.*, 'D3.2 Case study reports on constructive findings on the prerequisites of successful big data implementation in the transport sector', Jun. 2019.

[3] NOESIS, 'D2.2 big data implementation context in transport', Oct. 2018.

[4] M. Carpio, D. Collado, J. L. Guerrero, and I. Guinea, 'Transforming Transport - D4.3 – Smart Highways Release 1'. Jun-2018.

[5] J. Graf, M. Schygulla, D. Cobo, and P. Campillo, 'Transforming Transport - D5.3 – Connected Vehicles Pilots Release 1'. Jun-2018.

[6] Á. Pérez Bartolomé, J. M. Alonso Pérez, M. Rodríguez Plaza, and P. Flynn, 'Transforming Transport - D6.1 – Proactive rail infrastructures pilots design'. Mar-2017.

[7] F. Blanquer *et al.*, 'Transforming Transport - D7.1 – Port Pilots Design'. Mar-2017.

[8] J. Oguiza *et al.*, 'Transforming Transport - D8.1 – Smart Airport Turnaround'. 31-Mar-2017.

[9] J. Scholliers, M. Galende, M. Angeles Gallego, J. Laakso, J. Kanninen, and T. Korhonen, 'Transforming Transport - D9.1 - Integrated Urban Mobility'. Mar-2017.

[10] E. Zampou, C. Milioti, F. Katerina, A. Liapis, and I. Kelepouri, 'Transforming Transport - D10.1 - Dynamic Supply Chain Networks'. Mar-2017.

[11] P. Bohn, I. Deosdad, J. Escamilla, J. Franke, T. Jansen, and D. Mayer, 'Transforming Transport - D7.3 – Ports as Intelligent Logistics Hubs'. Jun-2018.

[12] K. Hee, T. Ivanov, R. V. Zicari, and R. Waldenfels, 'D1.3 big data Methodologies, Tools and Infrastructures', 2018. [Online]. Available: https://static1.squarespace.com/static/59f9cdc2692ebebde4c43010/t/5b6d4674032be489a442fa8b/1533888127770/20180716_D1.3_Big+data+methodologies%2C+tools+and+infrastructures_LeMO.pdf. [Accessed: 05-Aug-2019].

[13] T. Ivanov, S. Izberovic, and N. Korfiatis, 'The Heterogeneity Paradigm in big data Architectures', in *Artificial Intelligence: Concepts, Methodologies, Tools, and Applications*, IGI Global, 2017, pp. 485–511.

[14] C. Katrakazas, 'NOESIS - D2.1 big data and emerging transportation challenges'. 30-Jun-2018.

[15] L. Rocher, J. M. Hendrickx and Y.-A. de Montjoye, 'Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models' (2019) 10(3069) Nature Communications <https://doi.org/10.1038/s41467-019-10933-3> accessed 12 September 2019

[16] T. Teoh *et al.*, 'Deliverable 2.1 Report on Economic and Political Issues'. 31-August-2018. [Online]. Available: https://lemo-h2020.eu/s/20180831_D21_Report-on-Economic-and-Political-Issues_LeMO.pdf

[17] J. Debussche *et al.*, 'D2.3 Report on Ethical and Social Issues'. 31-August-2018. [Online]. Available: https://lemo-h2020.eu/s/20180829_D23_Report-on-Ethical-and-Social-Issues_LeMO.pdf

[18] S. Janković *et al.*, 'D2.3 big data in Transport Library'. 31-October-2018. [Online]. Available: https://drive.noesis-project.eu/index.php/s/B4hVPdQMLVxgU9j

[19] R. Pichler, 'D4.1 Summary to practitioners on Laws, Regulations, and Directives on Data Privacy, Security and Openness'. 27-February-2019. [Online]. Available: https://drive.noesis-project.eu/index.php/s/pj9G6XmMwLO6vAC

[20] University of Belgrade, Faculty of Transport and Traffic Engineering, 'D4.3 Handbook on Key Lesson learnt and Transferable Practices'. 28-June-2019. [Online]. Available: https://drive.noesis-project.eu/index.php/s/Uigt0ggLZtQ6WS2

[21] A. Monzón *et al.*, 'D2.2 – Analysis of Pilot Requirements for big data Use'. 1-January-2017. [Online]. Available: https://transformingtransport.eu/sites/default/files/2017-08/D2.2%20%E2%80%93%20Analysis%20of%20Pilot%20Requirements%20for%20Big%20Data%20Use.pdf

[22] J. Debussche et al., 'Deliverable D2.2 Report on Legal Issues'. 31-October-2018. [Online]. Available: https://lemo-h2020.eu/s/D22_Report-on-Legal-Issues_LeMO-FINAL.pdf

Appendix A: List of reference notations considered

Notation structures of reference are different according to the format "LIM-Aspect-Major issue (Initial)-Sub issue (Number)".

Table 34 List of reference notations considered

Code	Description
LIM-TM-DR-#	Limitations and barriers related to data resources in technical aspects
LIM-TM-DC-#	Limitations and barriers related to data complexity in technical aspects
LIM-TM-IS-#	Limitations and barriers related to limited infrastructures and systems in technical aspects
LIM-TM-AT-#	Limitations and barriers related to affecting technical solutions in technical aspects
LIM-POL-PU-#	Limitations and barriers related to public policy in policy aspects
LIM-POL-PR-#	Limitations and barriers related to privacy policy in policy aspects
LIM-LEG-DP-#	Limitations and barriers related to data privacy and protection in legal aspects
LIM-LEG-CBO-#	Limitations and barriers related to (cyber-) security and breach-related obligations in legal aspects
LIM-LEG-AP-#	Limitations and barriers related to anonymization and pseudonymisation in legal aspects
LIM-LEG-DIG-#	Limitations and barriers related to the supply of digital content and services in legal aspects
LIM-LEG-FF-#	Limitations and barriers related to the free flow of data in legal aspects
LIM-LEG-IP-#	Limitations and barriers related to intellectual property rights in legal aspects
LIM-LEG-OD-#	Limitations and barriers related to open data in legal aspects
LIM-LEG-DS-#	Limitations and barriers related to data sharing obligations on legal aspects
LIM-LEG-DO-#	Limitations and barriers related to data ownership in legal aspects
LIM-LEG-DSA-#	Limitations and barriers related to data sharing agreements in legal aspects
LIM-LEG-LI-#	Limitations and barriers related to liability in legal aspects
LIM-LEG-COM-#	Limitations and barriers related to competition in legal aspects
LIM-ES-TR-#	Limitations and barriers related to trust in ethical and social aspects
LIM-ES-SUR-#	Limitations and barriers related to surveillance in ethical and social aspects
LIM-ES-PRI-#	Limitations and barriers related to privacy in ethical and social aspects
LIM-ES-FW-#	Limitations and barriers related to free will in ethical and social aspects
LIM-ES-DO-#	Limitations and barriers related to data ownership in ethical and social aspects
LIM-ES-SD-#	Limitations and barriers related to social discrimination in ethical and social aspects
LIM-ES-ENV-#	Limitations and barriers related to environment in ethical and social aspects
LIM-ES-OT-#	Limitations and barriers related to others in ethical and social aspects
LIM-EP-IR-#	Limitations and barriers related to inaccurate results in economic and political aspects
LIM-EP-RAD-#	Limitations and barriers related to restricted access to data in economic and political aspects
LIM-EP-OLC-#	Limitations and barriers related to organizations lack capacity in economic and political aspects
LIM-EP-IU-#	Limitations and barriers related to big data industry underdeveloped in economic and political aspects
LIM-EP-UPR-#	Limitations and barriers related to uncertain political risk in economic and political aspects