LEVERAGING BIG DATA FOR MANAGING TRANSPORT OPERATIONS

Deliverable 3.2 **Case study reports on constructive findings on the prerequisites of successful big data implementation in the transport sector**

Tharsis Teoh[1], Bastiaan van Berne[1], Ivo Hindriks[1], Rut Waldenfels[2], Timo Eichhorn[2], Todor Ivanov[2], Minsung Hong[3], Rajendra Akerkar[3], Maruša Benkic[4], Jayant Sangwan[4], Rémy Russotto[4], Julien Debussche[5], Jasmien César[5]

[1]Panteia, [2]Goethe University Frankfurt
[3]Western Norway Research Institute
[4]CORTE, [5]Bird & Bird

June 2019

Work Package 3

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	30/6/2019 (M20)
Sent to coordinator	28/06/2019
No. of document	D3.2
Title	*Case study reports on constructive findings on the prerequisites of successful big data implementation in the transport sector*
Status & Version	*Final*
Work Package	3: LeMO case studies
Related Deliverables	D1.1, D1.2, D1.3, D2.1, D2.2, D2.3, D2.4, D3.1, D4.1, D4.2
Leading Partner	Panteia
Leading Authors	Tharsis Teoh, Panteia, Chapters 1,2,10 Rut Waldenfels & Timo Eichhorn, GUF, Chapter 3 Minsung Hong & Rajendra Akerkar, WNRI, Chapters 4, 5 Maruša Benkic, Jayant Sangwan & Rémy Russotto, CORTE, Chapters 6, 8 Ivo Hindriks & Bastiaan van Berne, Panteia, Chapter 7 Rut Waldenfels & Todor Ivanov, GUF, Chapter 9
Contributors	Julien Debussche & Jasmien César, B&B, Chapter 3 to 9 Michiel Muller, Panteia, Chapter 3 to 9
Reviewers	Jasmien César, B&B Julien Debussche, B&B Marusa Benkic, CORTE Jayant Sangwan, CORTE
Keywords	Case study; technological; political; economic; social; ethical; legal; environmental; rail; open data; real-time traffic management; optimized logistics; inland shipping; multimodal transport planner; mobility on-demand

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

The deliverable presents seven reports of the case studies conducted in Work Package 3 during Task 3.2. The case studies conducted are the following:

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

The methodology outlined in D3.1 "Case study methodology" was used as a template for each of the case studies. The template provided a consistent, but flexible approach to address the unique circumstances and learnings in each case study. It also leveraged the case study leaders' strengths in understanding the applications of big data technology in transport operations.

Besides developing a deep understanding of the big data technology and its business applications, the case studies also present an analysis of the issues that serve as 'opportunities' and 'barriers' to the implementation of big data, as well as the resulting outcomes of the implementation. These issues were analysed using the knowledge developed in Work Packages 1 and 2 of the LeMO project, from economic, political, social and ethical, legal, and environmental perspectives.

Case study 1 focused on the application of big data technologies in predictive maintenance in rail transport. The digitalization of train and rail infrastructure operations produces valuable data, which when analysed with the right big data technologies, can improve the prediction of potential train-related failures and the precision of maintenance schedules, thus improving overall safety and fleet availability. As rail transport forms the backbone of the transport strategy of many Member States in the EU, any promising improvement in railway operations will have a considerable positive impact on the EU as a whole.

Case study 2 explored the opportunities and challenges of successfully making big data available as open data. Open data is an important policy objective and strategy of the European Commission to support research and innovation in the field of transport. While it may seem simple to implement, making big data open entails many technological and organisational challenges, especially when economic incentives are lacking. Initiatives on open data are currently championed by transport authorities and organisations that rely on data, especially to keep up with customer demands and to provide innovative transport services through real-time and precise transport information.

Case study 3 explored the real-time traffic management which is one of important parts in Intelligent Transport Systems (ITS) that rely on big data analytics. ITS and the emerging field of C-ITS (Cooperative Intelligent Transport Systems) is effective in dealing with many challenges of managing road traffic, as well as in dealing with the emergence of connected and automated vehicles. While many transport management systems in Europe produce a wide variety and volume of data, there is still an untapped potential for the widespread use of big data techniques to improve traffic flows.

Case study 4 provides an in-depth description and analysis of the use of big data to support logistics processes in an era where customer and supply chain requirements are stricter and

more varied. The study is based on the very advanced and successful company Kepler51, which is based in the USA. The case study presents their big data solution that monitors, assesses, and forecasts transport risks and delays along the transport chain. However, their application, which focuses on non-vehicle data as a primary source, does serve to illustrate some of the potential benefits and risks facing transport carriers in Europe when considering the use of big data technology.

Case study 5 introduces a unique application of big data to coordinate the traffic of both the road and inland waterway network in the Netherlands. Coordination of inland waterways is an important topic in the Netherlands, especially since the network intersects with the road network at bridges. The major challenge faced was the political challenge to organise the collaboration between powerful governmental bodies and port authorities. But, once the initial challenge was met, the benefits of the programme were acknowledged, spurring replication in other parts of the Netherlands. While other countries in the EU may not have a similar application (i.e. the integration of the road and inland waterway network), the lessons learned in this case provide important clues as to the organisational challenges of multi-stakeholder large-scale project implementation.

Case study 6 presents the application of big data technology and analysis methods to provide a multimodal public transport information and route planning service to passengers. The challenge has been to develop a service that provides real-time information to passengers, while preserving the brand and identity of the individual public transport operators. This also offers an important opportunity for identifying an alternative to current non-European service providers that reduce the digital autonomy in Europe, thus leading to an imbalanced market. The service discussed in the case-study supports improved mobility and customer experience by providing effective route planning by combining open data from various sources. It also intends to further improve its product in the future by combining data collected from end-users (e.g. through data from fare cards), but this seems difficult on account of privacy concerns expressed by end-users. The case serves as an important source of knowledge about the issues that many cities are facing in terms of both improving the multimodal transport information services as well as route planning, especially in the context of disruptive and innovative transport services (e.g. scooter sharing).

Case study 7 presents the challenging and disruptive on-demand urban mobility services sector. A disruptive technology company working in the heavily regulated urban transport sector needs to be innovative in order to tackle the challenges of policy, a small customer base, and the competition with established firms. Furthermore, the technologies supporting the activity have to be very reliable and attractive so that new customers easily become regular customers. The case presents a spin-off from an established company aiming to cover the gaps in their existing services. Big data, especially open data, empowers the start-up to provide niche services that fill the gaps in the existing public transport system.

The case studies provide a realistic view of the opportunities and benefits provided by big data in the specific domains. However, they also bring to light the very many challenges that must be overcome for a successful and sustainable improvement of transport sector using big data. The issues highlighted must therefore be carefully addressed via policy at the right governmental levels and via further research to improve the technical application of the technology. The LeMO consortium will rely on the results of these case studies in the next phase of the project in Work Package 4 to conduct the horizontal analysis and road-mapping exercises.

Table of contents

List of Figures

List of Tables

Glossary

Abbreviation	Expression
AFC	Automated Fare Collection
AI	Artificial Intelligence
AIS	Automatic Identification System
APC	Automated Passenger Counter
API	Application Programming Interface
ATP	Automatic Train Protection System
AWS	Amazon Web Services
B2B	Business-to-business
B2C	Business-to-consumer
B2G	Business-to-government
BDT	Big data technology
BGV	Blauwe Golf Verbindend project (in English, "Blue Wave Connected")
BLIS	Binnenvaart Ligplaats Informatie Systeem (in English, "Inland Waterway Berth Information System")
BMC	Business Model Canvas
BMS	Bridge Management System
CC	Creative Commons
CCTV	Closed-circuit television
C-ITS	Cooperative Intelligent Transport Systems
CO2	Carbon dioxide
CVSA	Commercial Vehicle Safety Alliance
DTLF	Digital Transport and Logistics Forum

EC	European Commission
EFTA	European Free Trade Association
ERTMS	European Railway Traffic Management System
ETCS	European Train Control System
G2B	Government-to-business
G2C	Government-to-consumer
G2G	Government-to-government
GDPR	General Data Protection Regulation
GHG	Greenhouse gas
GPS	Global Positioning System
GSM-R	Global System for Mobile Communications – Railway
IoT	Internet of Things
IP	Internet Protocol
IWW	Inland Waterway
JSON	JavaScript Object Notation, a lightweight data-interchange format
LeMO	Leveraging Big Data for Managing Transport Operations
MOD	Mobility On-demand
MQTT	Message Queuing Telemetry Transport, an ISO standard (ISO/IEC PRF 20922) publish-subscribe-based messaging protocol.
NAFTA	North American Free Trade Agreement
NAP	National Access Point
NDW	National Data Warehouse
NOAA	National Oceanic and Atmospheric Administration (US)
NPRA	Norwegian Public Roads Administration

NRDB	National Road Database
NTS	National Traffic Management System
OEM	Original Equipment Manufacturer
OGP	Open Government Partnership
P2P	Peer-to-peer
PSI	Public Services Information
PTO	Public transport operators
PTW	Powered two-wheel driver
RIS	River Information System
RTTI	Real-time traffic information
SafeSeaNet	A vessel traffic monitoring and information system in the EU
SCADA	Supervisory Control and Data Acquisition
SCM	Supply Chain Management
SESAR	Single European Sky ATM Research
SMP	Smart Mobility Planner
TCP/IP	Transmission Control Protocol/Internet Protocol
TDT	Transport Department of Tallinn
TMC	Traffic Management Centre
TMS	Traffic Management Systems
USDOT	US Department of Transportation
VRU	Vulnerable Road User
WP	Work package

1 Introduction

1.1 Abstract

The deliverable presents seven reports of the case studies conducted in Work Package 3 during Task 3.2. The case studies conducted include:

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

The methodology outlined in D3.1 "Case study methodology" was used as a template to be followed for each of the case studies. The template provided a consistent, but flexible approach to address the unique circumstances and learnings in each case study. It also leveraged the case study leaders' strengths in understanding the applications of big data technology in transport operations.

Besides developing a deep understanding of the technology and business applications of big data, the case studies also present an analysis of the issues that serve as opportunities and barriers to the implementation of big data as well as the resulting outcomes of the implementation. These were analysed using the knowledge of the issues developed in Work Packages 1 and 2 of the LeMO project, from economic, political, social and ethical, legal, and environmental perspectives.

1.2 Purpose of the document

The document is a consolidation of the case study reports as developed by the individual case study leaders, members of the LeMO project consortium. The main purpose is to support the thematic horizontal analysis and road-mapping exercises to be carried out in Work Package 4. The structure adopted in the case study reports mirrors the literature review and analysis conducted in WP2 of the impact of big data applications.

1.3 Target audience

The results of the research are especially interesting for people working for organisations in the public sector (e.g. politicians, policy makers, policy consultants etc.) and in the private sector (e.g. managers, directors, and consultants).

More specifically, the target audience for this deliverable includes:

- European Commission
- Public and private transport organisations
- Horizon 2020 projects and related transport projects (cf. clustering activities)
- Partners, Case providers and Advisory & Reference Group in the LeMO project

2 Presentation of the case study reports

Deliverable 3.2 consolidates the reports of the case studies conducted as part of Task 3.2. Each case study report follows a set of guidelines and a template developed in Task 3.1 "Case study methodology". This chapter presents the main research questions and a general template of how the individual case studies are presented.

The case study reports are organised in Chapters 3 to 9 as follows:

- Chapter 3: Case study 1 "Railway transport"
- Chapter 4: Case study 2 "Open data and the transport sector"
- Chapter 5: Case study 3 "Real-time traffic management"
- Chapter 6: Case study 4 "Logistics and consumer preferences"
- Chapter 7: Case study 5 "Smart inland shipping"
- Chapter 8: Case study 6 "Optimised transport & improved customer service"
- Chapter 9: Case study 7 "Big data and intelligent transport systems"

Each case study follows a similar template, though the unique properties of each case study might necessitate modifications to the report structure.

Chapter 10 concludes the deliverable with a preliminary analysis of the results. It highlights several main themes identified across the case studies. Further research avenues will be recommended, which should be taken up in WP4.

2.1 Objectives and methodology

The existing knowledge of the (potential) use of big data in transport operations is still in its infancy. Different private and public organisations have taken up the challenge to incrementally develop new and optimise existing products and services using the state-of-the-art technology. The case study reports presented in this deliverable are reports on how the respective organisations are currently using big data technology (BDT), as well as their experiences in the implementation, especially when dealing with the technical and business opportunities and challenges, as well as the economic, political, social, ethical, legal, and environmental issues.

The sources of information for each case study differ depending on the availability of information. In general, semi-structured interviews were carried out with relevant members of the organisation. Additional literature reviews using a variety of peer-reviewed and non-peer-reviewed material was used to develop further the case study context. The case study reports have been reviewed by the member of the organisation interviewed as part of the verification process.

2.2 Case study report structure

Each case study report covers a separate theme, based on interviews with a principal organisation. The structure of the case study report is presented below.

2.2.1 CASE STUDY OVERVIEW

This section presents an overview of the case study and the subject. It is meant to give a broad context for interpreting the business process and issues described further in other sections.

Organisation/Stakeholders	Name of the organisation of the case study, public/private and country
Sector/Mode	Transport sector and mode of the organisation
Case study motto	Short phrase with the main idea of the case study
Executive summary	Brief description of the case study, explaining its rationale and stressing the main points

2.2.2 TECHNICAL PERSPECTIVE

In this section, the focus is on data and BDT considered and used by the subject. As BDT and applications are itself still in its infancy, there are limitations that serve as hard constraints on the viability of the use. This section presents the context of the BDT applications from the perspective of the technology and infrastructure needed.

Data sources/Uses	Description of the main characteristics of the data sources/uses
Data flows	Description of the data as a resource for the case study, examining where data originate, where they flow and where are processed
Relevant big data policies	Big data policies relevant for the case study and brief explanation of their importance
Main technological challenges and issues	Brief explanation of the main technological difficulties in the area of case study including expectations from the use of big data Also other aspects such as hiring strategies, analytics landscape (descriptive/predictive)
Big data dimensions and assessment	Assessment of the importance of the volume, velocity, variety, veracity, and other big data dimensions for the case study

2.2.3 BUSINESS MODEL

In order to understand a business model in a straightforward and structure way, each case study will use the business model canvas as shown in Figure 1. This snapshot is a static view of the business model, and serves to highlight especially the innovations introduced by the BDT.

Business model	The core strategy and overview of the business plan of the firm in its integration of big data is presented in the business model canvas.
Business processes	Major business processes involved in the case study
Relation to big data initiatives	Big data initiatives related to the case study
Illustrative user stories	Brief and relevant set of user stories that capture the user point of view of the case study

<table>
<tr><th>Key partners</th><th>Key activities</th><th>Value proposition</th><th>Customer relationships</th><th>Customer segments</th></tr>
<tr><td rowspan="3">Who are the firm’s partners?</td><td>What are the key activities, such that the firm operates successfully?</td><td>What value does the firm deliver to the customer?</td><td>What type of relationship is established between the firm and the customer?</td><td rowspan="3">For whom is the firm creating value?</td></tr>
<tr><th>Key resources</th><th>Internalization of costs and benefits</th><th>Channels</th></tr>
<tr><td>What does the firm need to create value?</td><td rowspan="3">How are costs and benefits from firm’s business activities internalized?</td><td>How does the firm reach the customer?</td></tr>
<tr><th colspan="2">Cost structure</th><th colspan="2">Revenue streams</th></tr>
<tr><td colspan="2">What are the costs associated to the business model?</td><td colspan="2">For what value are the customers paying and how are they charged?</td></tr>
</table>

Figure 1 Business model canvas[1,2]

2.2.4 ANALYSIS OF SOCIETAL ASPECTS

An important aspect of policy analysis is to understand the societal impacts, whether positive or negative. In this section, the following aspects are considered, primarily using the deliverables from WP2 as background knowledge:

- economic and political, based on D2.1 “Report on economic and political issues”
- social and ethical, based on D2.3 “Report on ethical and social issues”
- legal, based on D2.2 “Report on legal issues”
- environmental, based on D2.4 “Report on trade-off from the use of big data in transport”

Interviewees are asked for their opinions on the list of issues discovered in the various WP2 reports. Where relevant, issues from additional literature are added and discussed in the context of the case study.

Economic and political	Description of positive and negative economic and political issues that were found in the case studies
Social and Ethical	Description of positive and negative social and ethical issues that were found in the case studies
Legal	Description of positive and negative legal issues that were found in the case studies

1 Rodrigues, M., Zampou, E., Zeimpekis, V., Stathacopoulos, A., Teoh, T., & Ayfantopoulou, G. (2018). Cooperative Models for Addressing Urban Freight Challenges: The NOVELOG and U-TURN Approaches,. In E. Taniguchi & R. G. Thompson (Eds.), City Logistics 3: Towards Sustainable and Liveable Cities (pp. 215–233). ISTE.

2 TURBLOG. (2011). FP7 TURBLOG Deliverable 2 Business Concepts and Models for urban logistics.

Environmental	Description of positive and negative environmental issues that were found in the case studies

2.2.5 ASPECTS RELATED TO OPPORTUNITIES, BARRIERS AND LIMITATIONS

In this section, various opportunities, barriers and limitations are discussed, based on the technical, business, and societal impact findings.

Opportunities	Description of the opportunities that were found in the case studies
Barriers and limitations	Description of the barriers and limitations that were found in the case studies and constructive suggestions for diminishing barriers and limitations

2.2.6 SOURCES OF INFORMATION

The sources of information employed in the elaboration of the case study are presented at the end of each case study report. This covers both the interviewees, as well as the literature referenced.

3 Case study 1: "Railway transport"

The case study on "Railway transport" focused on the role of big data in predictive maintenance in the rail transport sector using the example of the Siemens Mobility Application Suite Railigent®.

3.1 CASE STUDY OVERVIEW

3.1.1 Organisation/Stakeholders

A much discussed and promising new development in the course of digitisation in the rail transport sector is the so-called predictive maintenance. This case study examines the concrete example of a predictive maintenance service mobility platform: Siemens Railigent (Siemens AG, 2019a). Railigent is the rail application suite from Siemens Mobility GmbH. It is powered by MindSphere - the open IoT operating system from Siemens [4].

Siemens Mobility GmbH is a subsidiary of the Siemens Group in Germany that has existed since 2014. Siemens Mobility GmbH also manufactures rail vehicles such as the ICE4, the Velaro and the Rhein-Ruhr-Express. Siemens Mobility GmbH employs a total of over 28,000 people.

With Railigent, Siemens Mobility wants to offer its customers 100% system availability of infrastructure and rolling stock in rail transport. This means that both the infrastructure and the corresponding rail transport vehicles are ready for operation and fully functional at all times according to plan. The Railigent team has no influence whatsoever on all other processes, such as personnel planning or boarding and alighting procedures. By increasing availability, rail operators are in turn able to guarantee their customers punctuality and guarantee the reliability of the timetable for both passenger and freight transport. The Railigent team uses Data Analytics methods to achieve this goal.

With Railigent, Siemens operates in over 30 countries worldwide, including China and the USA. Railigent also supports rail transport operators in almost every country in Europe, for example in Germany, Austria, England, Spain, Finland, Norway, Poland, Italy and France. Customers are private or state transport companies or train and infrastructure owners or manufacturers. In total, the company has slightly more than 60 customers worldwide. Railigent® open ecosystem was successfully launched at Innotrans 2018 with currently 15 partners (state 04/2019). Thus, the digital service offering is enriched with other rail applications accessible via Railigent®.

Table 1 gives an overview of the stakeholders generally involved in predictive maintenance in the rail sector (Zaldivar, 2017). The stakeholders involved can roughly be categorised into manufactures, operators, customers or users and other regulating parties.

Table 1 Stakeholders in the field of predictive maintenance in the rail transport sector

Stakeholder	Interest / characteristics
Manufacturers of rolling stock and infrastructure	Build rolling stock and infrastructure, want to sell them. Can also upgrade infrastructure and rolling stock with sensors for better data collection.

Owner of rolling stock and infrastructure	Manage and own the rolling stock and infrastructure, own also collected data of stock and infrastructure. Provide infrastructure or rolling stock to the operators. Might be operators themselves.
Operator of rolling stock and infrastructure	Operate and manage train journeys and services to offer the transport service to the public or to private companies, monitor operations and collect data.
Predictive maintenance service provider	Offer predictive maintenance services, need access to operational data to offer these services.
Railway regulatory body	Governmental organisation regulating the rail transport sector, may also regulate the relevant data flows and uses.
Railway industry organisation	Organisations supporting and advising companies and operators in the rail transport sector.
Passengers and logistic companies	Use rail services, are interested to protect their personal/ business data, want reliable transport services.

3.1.2 Sector/Mode

This case study focuses on rail transport. The topic of predictive maintenance is relevant for all sub-sectors of rail transport, e.g. passenger and logistics transport, regional and long-distance train services.

3.1.3 Case study motto

This case study investigates the new maintenance strategy of predictive maintenance and its implications in the rail sector using the example of the Railigent platform of Siemens Mobility.

3.1.4 Executive summary

In this case study, the concept of predictive maintenance, its relation to big data and its concrete implementations are explained. The business model of service provider is discussed and topic related issues of social, legal and environmental nature are explained.

In a simplified picture one can differentiate three major maintenance strategies of any kind of infrastructure (Arbour, 2019; Buurmeijer et al., 2018):

1. **Condition-based or reactive maintenance:** This type of maintenance is carried out when a problem, such as the failure of a component or similar, has already occurred. Repairs are not planned. This type of maintenance is not cost-intensive in itself, but involves the great danger of causing major delays in operations and thus high costs.
2. **Planned or preventive maintenance:** Planned maintenance is carried out at regular intervals. This type of maintenance is more cost-intensive than reactive maintenance, since inspections take place more frequently and need to be carried out by trained

personnel. This means that breakdowns can often be prevented and repairs carried out in time. For the personnel carrying out the work, this type of maintenance bears the risk of encouraging inattention, as many inspections are simply repeated without any changes being made. This in turn increases again the risk of errors.

3. **Predictive maintenance:** Predictive maintenance aims to predict infrastructure failures before they occur, enabling them to be repaired at the right time. Unnecessary routine checks become obsolete without the feared failure occurring. However, often higher investment costs in the infrastructure are necessary.

Predictive maintenance is based on data collected on equipment during operation. The so-called Internet of Things (IoT), i.e. the accessible information of all possible sensors of things, in our case trains and their infrastructure, enables the acquisition and evaluation of millions of data points. They are used to identify maintenance problems in real time and solve them promptly. This means that repairs can be planned on time, i.e. neither too early nor too late, and trains do not have to be taken out of service unexpectedly for emergencies or unnecessary routine maintenance. Figure 2 illustrates what constitutes the "right" maintenance timing.

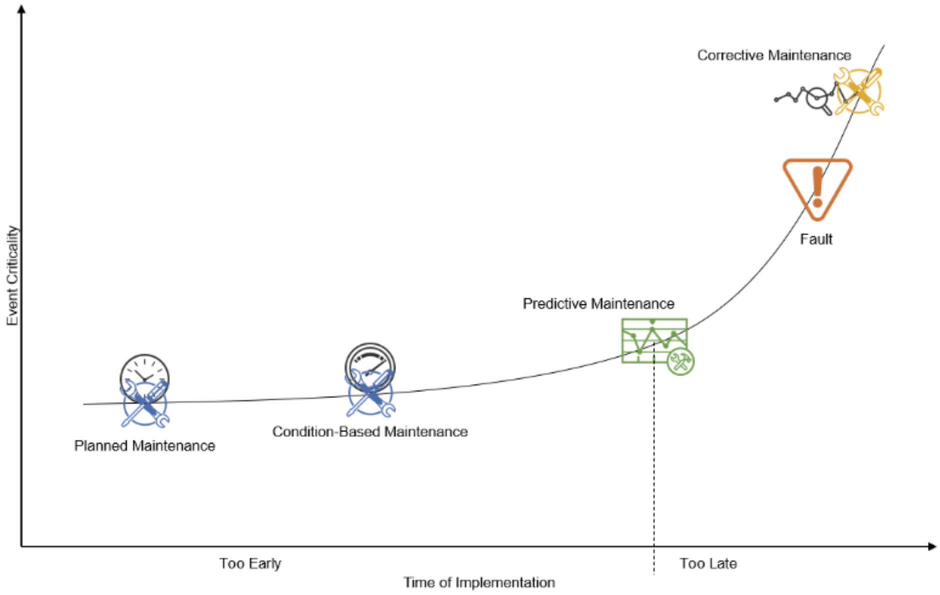

Figure 2 The “right” maintenance timing for rail. (Lugara, 2018)

Predictive maintenance is becoming increasingly important for regional railway companies and freight railway operators in particular as a result of increased competition. This can help maintain competitiveness. According to a study by the McKinsey Group (Stern et al., 2017), the use of such analyses is estimated to increase efficiency by 10-15% in rail transport. However, the additional necessary investments in modern technology limit the achievable cost savings to a maximum of 10%, which is currently not yet considered significant enough. Jobs and responsibilities in the field of maintenance will change drastically, as many maintenance tasks will be automated.

Train and infrastructure manufacturers are also increasingly becoming service providers, in particular by offering Train-as-a-Service models that guarantee train availability. In contrast to other means of transport, such as motor vehicles, train fleets have a very long service life. In order to amortize the high initial investment, the operation of a vehicle should be guaranteed for as long as possible. For this reason, too, new technologies that enable predictive maintenance, reduce operating costs and extend the life of a fleet offer enormous financial benefits. On the other hand, some older trains have to be retrofitted with more modern sensor technology in order to provide sufficient data for predictive analysis, which in turn requires investment. Figure 3 shows the estimated increase in efficiency through the combination of conventional and predictive maintenance in rail transport.

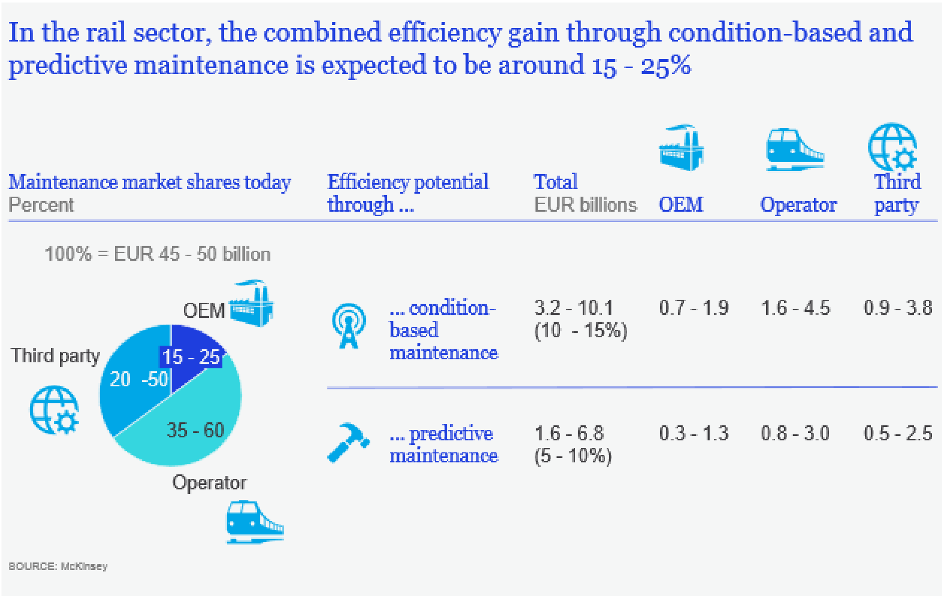

Figure 3 Efficiency gain through combination of condition-based and predictive maintenance in rail sector (Stern et al., 2017)

3.2 TECHNICAL PERSPECTIVE

3.2.1 Data sources/Uses

There are two main data sources for rail applications: rolling stock and infrastructure. Important data comes, for example, from the sensors installed in the locomotives, the wagons, the signal boxes, the switches and the signalling systems.

Roughly speaking, there are 3 categories of valuable data in the field of rail transport:

- **Process data**: This refers to continuous data that is produced continuously during operation. In concrete terms, this means, for example, information on current speeds or voltage curves on the locomotives. This data is produced in seconds and called up via a central interface. Although the individual data points are very small, large amounts of data are produced over time. In order to optimize this, often only a so-called delta observation is made. This means that data is only sent or recorded if there is a significant change in the course of time.
- **Diagnostic messages**: These are messages that are transmitted to the system in the event of a failure that has already occurred. This is, for example, the information that the vehicle engine is overheating or that the air conditioning system has failed.
- **Log messages**: Messages concerning the use of a component, e.g. outdoor lighting is on or off. Log messages only occur when the actual status changes.

In order to build databases with useful information for predictions, all possible measurable variables and their measuring ranges or measuring technique are considered. Some examples are presented in *Table 2*.

Table 2 Relevant data for rail traffic (Lugara, 2018)

Category	Components
Process data	Axles (stress, load, forces), brakes, wheels, Voltages from current sensors, etc.
Diagnostic messages	Door systems, bogies, filters, water- and air pressure sensors, acoustic sensors to detect vibrations, Temperature sensors, Oil compounds: measuring the compounds in the lubrication of sensitive parts, etc.
Log messages	Door systems, Pantographs, vibrations: Shock pulse measurement, etc.

All sensors can be connected via networks of spatially distributed intelligent nodes. A node represents a single sensor, power supply, microcontroller, and IP (Internet Protocol) data transmitter that allows the device to use the TCP/IP protocol.

To avoid duplication of measurements from different sensors, the nodes must be positioned along the train's subsystems. Furthermore, due to the radio range of the sensors, it is essential to uniquely identify each sensor and connect it to its train. Otherwise, due to the radio range of the sensors, data from one train could be collected and assigned to another train, if the two trains are close to a line or station.

Some sensors in the vehicles are connected to the system by their diagnostic messages, but the continuously recorded data is not shared. The subsequent connection of such previously isolated sensors is currently taking place in many fleets.

Trains and infrastructure have comparatively few breakdowns or faults over the operating period, and the systems are very reliable in comparison to the overall collected data of the whole fleet and infrastructure. This has analytical implications. For example, this leads to internal correlations. If one searches for correlations in the data without concrete research questions or problems, this often leads to the identification of false correlations. For example, for a traction motor failure, an event set of 6 to 8 failures must be sufficient to train the corresponding predictive models on the basis of this data.

The sensor technology installed in many current trains, locomotives and the corresponding infrastructure is often more than 15 years old. Acoustic sensors in particular and cameras for video analysis of the components are retrofitted. Despite the old sensor technology, it is often possible to make reliable predictions with the existing data sources. Here it is always a question of weighing the costs of a fleet-wide retrofit against the savings potential of predictive maintenance.

3.2.2 Data flows

Railigent is built on top of a Siemens MindSphere architecture. MindSphere is a cloud-based, open operating system for IoT. The different data sources provide the information to the MindSphere data lake by a secured connection to a cloud service. On top of the data lake, several applications can be executed. The applications are built to analyse the collected data (Siemens AG, 2018).

Figure 4 MindSphere schematic architecture

The data originating in the described sources of the previous section is managed mainly via Amazon Webservices (AWS), so it is cloud-based. Railigent has to handle a large number of data types and formats, like diagnostic messages, sensor data, work orders, spare part movements, and images.

The received data comes in all sorts of legacy formats, most of them are batch formats. These files are decrypted and then catalogued. It is annotated with specific information to enable quick retrieval of the data, and also to ensure it can be attributed to the right fleet and the right customer. Then a generic JSON file is created which is stored in Railigents data lake. For streaming data, MQTT as a transfer protocol is mostly used.

The analyses offered should help to quickly identify the cause of an error or an imminent failure. In the various areas of rail traffic and at the various devices, such as turnout drives or level crossings, data analysis should ensure increased availability and predict possible faults. Railigent not only offers descriptive analytics, but also predictive and prescriptive analytics (Hee et al., 2018).

Railigent is built on technologies from Mindsphere, enriched with rail-specific applications like data models, rail specific format translators and specific data analytics models. MindSphere and Railigent share a lot of core functions, especially in the way the data connectivity and data handling is implemented and how the security of the system is ensured. The key reason to use the same technology is that it is essential to their customers to have a secure and reliable platform. And the key differentiator they provide is generating the insight. Therefore, the base platform functionalities are not different, and there is no need for developing them again.

The foundation is a data lake in the cloud, AWS, in which the data is stored in a loosely coupled format. The data gets ingested to the data lake in batch or stream, depending on the source. Even during the data ingest, preliminary analytic models are applied to validate and augment the data. As much as possible, Railigent relies on platform services from AWS to build the applications. The offered applications consist of micro services which are bundled in a common user interface framework (Zicari, 2018).

The data analytics is either done in a sand box, when the model is still in development, or in the full platform. As languages mostly Python, pySpark and R are used, but also other technologies are used when needed within the specific use case.

Most of the data analytics in Railigent is based on machine learning or deep learning. This can be classifiers to identify components which are already showing distress, or it can be prediction algorithms to identify the remaining useful life of a component. Most of the machine learning is supervised learning, so that a function is learned from an input type to an output type using training data that have both input and output values. But there are also cases where unsupervised learning techniques are implemented, in these, the learning process seeks to learn a structure in the absence of an identified output or feedback of the given data (Sammut and Webb, 2010). Most algorithms can be applied for different use cases, since they follow a general logic, but need to be adjusted to the actual physical environment of the stock and infrastructure.

There are essentially three levels of data analysis, which can be distinguished into descriptive, predictive and prescriptive analysis:

- descriptive analysis to present the current situation,
- prescriptive analysis that increase performance during operation in real time,
- predictive analysis, which includes simulation and prediction of train operation.

The descriptive analysis and the real-time prescriptive analysis are based on historical and real-time data of the fleet and the infrastructure. Descriptive analysis is used for simplifying a great amount of data for human interpretable information. So, at first the data is analysed and presented, then actions for a performance increase are suggested. For the third level, in addition to this historical data, further context and metadata are required, which is implemented with the help of machine and deep learning algorithms.

For the analytics side, Railigent uses all types of analytics libraries, but also mathematical approaches newly developed by Siemens. Especially for the industrial data area, new mathematical approaches are often required and such approaches were then integrated into Railigent.

3.2.3 Main technological challenges and issues

Bandyopadhyay et al. (2011) identified the following challenges and open issues across several subsection in the area of IoT. The key challenges are which have to be addressed are:

- **Network Foundation**: The current internet architecture is limited with regard to mobility, availability, manageability and scalability.
- **Security, Privacy and Trust**: The architecture has to be secured in terms of a secure design, identification and protection of arbitrary attacks and abuse, and identification and protection of malicious software. Regarding user privacy, the main focus is on control over personal information, need of privacy improvement technologies, and standards / methodologies for identity management. In the domain of trues simple and reasonable exchange of protected and sensitive data.
- **Managing heterogeneity**: Handle different applications, environments, devices, and communication standards.

Railigent addresses most of the challenges by their architecture i.e. by using a cloud infrastructure, secured transmission and storage of the data, and a wide range of supported protocols to connect different data sources.

3.2.4 Big data dimensions and assessment

The classification refers to all data collected for the analysis, including two data sources and three different categories described in the previous section. It is classified as a whole because all the data is collected in one data lake.

Table 3 Big data dimensions for the railway data lake

Dimension	Classification	Description
Variety	Very high	Various sensor inputs with different formats. Operating with transnational rail transport is challenging, there are different systems regarding: infrastructure (e.g. track width), electricity systems, legal regulations and also often software wise. The data heterogeneity is strong.
Velocity	Low to very high	Protocol information which is slowly renewed to sensors sending data points every milliseconds.
Volume	High	Stock and infrastructure are producing a huge amount of date over time.
Veracity	Low to moderate	Sensors might produce data with a certain failure rate. Also data is collected from several data sources and data providers.

3.3 BUSINESS MODEL

The business model of the Railigent is presented in Figure 5.

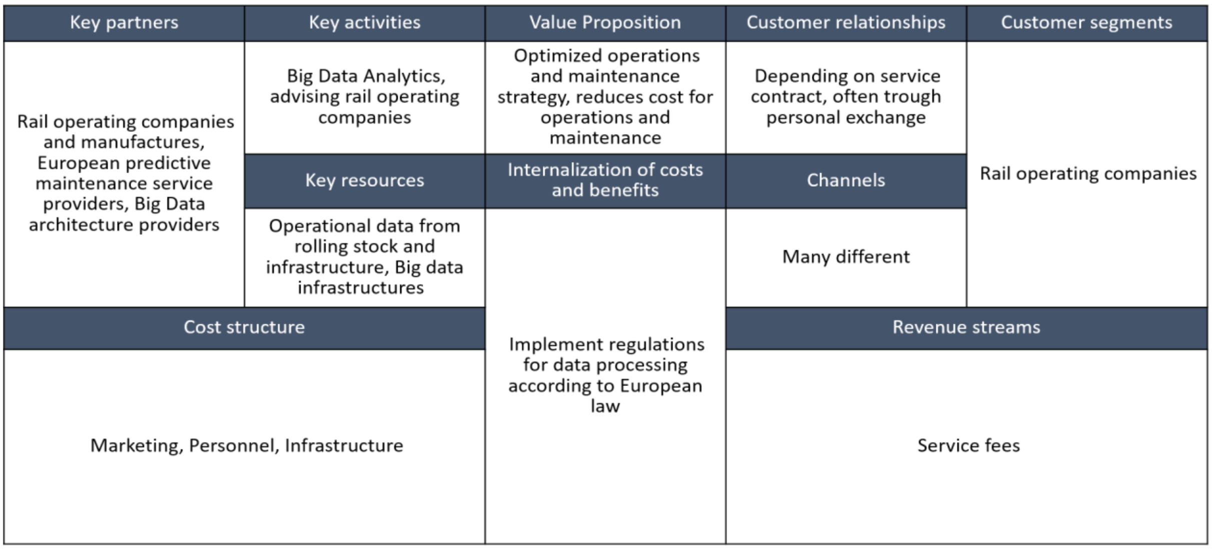

Figure 5 Business Model Canvas for predictive maintenance service providers

In the following each part of the canvas is briefly described:

- *Value proposition*. The Railigent services provide optimized operations and maintenance strategy and this reduces costs for operations and maintenance of the customers.
- *Customer segment*. The customer segment targeted are rail operating companies.
- *Customer relationship.* Depending on service contract different channels are used. Often the interaction is realised by personnel exchange.
- *Channels.* Many different channels are used to deliver the value proposition to the customers.
- *Key activities.* Big data analytics and advising rail operating companies are the key activities to produce and deliver the services provided by Railigent.
- *Key resources.* The key resources are operational data from rolling stock and infrastructure as well as the Big data infrastructures.
- *Key partners.* Rail operating companies and manufactures, European predictive maintenance service providers, Big data architecture providers.
- *Cost structure.* The major cost categories are marketing, personnel, and infrastructure expenditures.
- *Revenue streams.* The revenue is received by service fees of the customers.
- *Internalisation of costs and benefits.* Implement regulations for data processing according European law.

3.3.1 Business processes

In general, predictive maintenance services are based on operational data, which need to be made accessible for the service provider. In most cases in Europe service providers have individual contractual agreements with their customers defining the use and access to the relevant data. Depending on the needs of the customer, different levels of analytics are implemented. The details of the analysis and relevant services are then offered to the customer (see Figure 6).

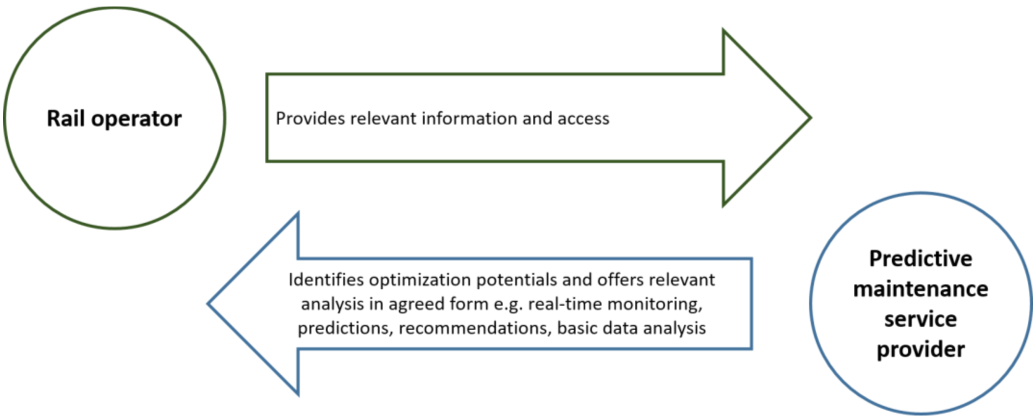

Figure 6 Simplified scheme of the business processes of predictive maintenance

Railigent is a cloud-based service, designed to help rail operators and rail asset owners, to improve fleet availability and improve operations. For example, by enabling intelligent data gathering, monitoring, and analysis for predictive maintenance in the rail transport industry.

Depending on the rail operators' own expertise and capabilities, the Railigent team is given various service areas. These range from a full-service maintenance package, in which all maintenance and servicing is carried out and coordinated by Siemens, to smaller analysis processes. This applies above all to companies in the rail sector, which themselves already have a well-developed infrastructure and the necessary expertise in advanced data analysis. Long-term contracts are often concluded for cooperation up to the year 2060, for example.

For each customer, weaknesses in the process flows that require optimization are identified and a solution or improvement is worked out through the use of data analytics. The challenge here that often very local and individual problems need to be solved. However, the solutions must at least be transferable to a certain degree in order for a developed algorithm to offer added value. Possible services for customers are, for example:

- Visualization of vehicle condition and location,
- Prediction of component failures for transmissions, bearings, traction motors, doors and power transformers,
- Analysis of the fault conditions of the European Train Control System (ETCS),
- Prediction of point machine errors and throughput analysis for the railway network.

3.3.2 Relation to big data initiatives

While there were no big data initiatives specific to predictive maintenance, 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.

3.3.3 Illustrative user stories

Some of the use cases that illustrate the potential benefits of predictive maintenance using big data are presented below.

Example 1 (Siemens AG, 2019b) - Long term analysis of rolling stock helps to deal with short-term analysis of anomalies

A locomotive that is on its way to a leasing operator MRCE in Poland is monitored from a rail support centre of Siemens Mobility GmbH in Erlangen. The team are working to predict potential faults affecting the rolling stock and to take action to prevent them. A warning occurs in one of the windows, indicating that the engine temperature of the locomotive is unusually high. They look at the history of the temperature data of this specific locomotive and also compare it to data from the fleet as a whole. The experts talk with the customer to determine how urgent the need for action is, and then take the most appropriate steps. Apparently, the assembly needs to be changed soon, but there is no immediate risk of component failure or of the engine overheating. Prescriptive identification made it possible to align the predicted maintenance with the operator’s action plan.

Example 2 (Siemens AG, 2019b, 2016) – Increased fleet availability

The Velaro E fleet in Spain was optimized using predictive maintenance services of Siemens Mobility. The 26 trains, including the engine and gearbox mounts, were equipped with monitoring systems. High operating speeds of up to 300 km/h mean that mechanical wear and tear is unavoidable, making predictive maintenance especially useful. The failure and delay rate of the trains was heavily decreased. One Velaro E departed with a delay of more than 15 minutes from a total of 2,300 journey, which equals rate of 0.004 percent. Due to that, on the route between Madrid and Barcelona, the ratio of air to train passengers, which was 80:20 as recently as 2008, has now been almost completely reversed.

Examples of monitoring of technical parts and the corresponding implementations (Siemens AG, 2019a):

- **Condition monitoring of ball bearings.** Remote diagnosis of possible bearing failures by means of diagnostic algorithms reduces the number of unexpected operational interruptions. Axle bearing faults and flat points are also taken into account, which extends the maintenance intervals.
- **Automatic video detection of track faults**. Detection of incipient faults in the railway infrastructure. The data generated by the video sensors are transmitted to the cloud and evaluated using algorithms for automatic image recognition.
- **Monitoring of special train couplings.** Remote monitoring enables the condition of deformation tubes and shock absorbers of Scharfenberg couplers to be recorded. The result is early detection of impacts and defective shock absorbers, thus reducing unscheduled downtimes.

3.4 ANALYSIS OF SOCIETAL ASPECTS

3.4.1 Economic and political

According to our interview partners, within Europe there is a tendency towards cooperation within the field of predictive maintenance in rail transport. This is due to the fact that the individual players tend to see competition outside Europe, for example in China.

In addition, there is an industry-wide departure from the standard supplier model, in which the players make strict time and service agreements. On the contrary, due to the complexity of this young business segment, there is a greater willingness for flexible exchange.

The McKinsey study (Stern et al., 2017) gives the following examples for developments in the rail sector in Europe due to digitalization and predictive maintenance:

- Stronger cooperation between rail operators and rolling stock manufactures. For example, SNCF (French railway) and Alstom cooperate since 2016 to design the next generation of the French high-speed train TGV. 4;
- Rail operators will outsource maintenance to service providers, who are often rolling stock manufactures. National Express and Abellio operate the new fleet of regional trains called “Rhein-Ruhr-Express” in the metropolitan areas of Nordrhein-Westfalen in Germany, Siemens Mobility was contracted to build and maintain the fleet of 82 trains over the next 32 years (starting in 2018). Siemens Mobility also maintains the 26 Velaro E high-speed trains for service between Barcelona and Madrid for the next 14 years;

- New cooperation between rail operators and software companies. For example, German cargo rail operator DB Cargo is now collaborating with General Electric to equip 250 of DB Cargo's locomotives with digital solutions from GE. The Italian rail operator Trenitalia cooperates with SAP to support its digital transformation;
- Many different players involved in the new maintenance possibilities: rail operators (e.g., Trenitalia, Deutsche Bahn), the rolling stock Manufactures (e.g., Siemens, Bombardier, Alstom), component suppliers (e.g., Knorr-Bremse), IT platform providers (e.g., Microsoft, SAP, IBM, GE), and analytics or sensor technology start-ups (e.g., Konux, Predikto, C3IoT). Rail operators and rolling stock manufactures are in the defining position, since they own a significant share of the data generated by the fleet, like operations data, track data, or data generated by sensors on the trains.

Figure 7 gives an overview of the current state of maintenance strategies in different European countries and ongoing related projects (Buurmeijer et al., 2018).

Country	Rail infra or rolling stock projects	Maintenance concept	Specialties
UK	Intelligent infrastructure and linear asset management (Network Rail) Various rolling stock online monitoring applications (Alstom, Hitachi)	Condition based	> 40,00 rail infra assets on line by Network Rail Company owned condition based maintenance platforms Alstom Hub and Hitachi Lumada
Germany	Monitoring of train bogies (DB) Monitoring and diagnostics of switches (DB Netz) Monitoring of elevators and lifts at stations (DB Station & services)	Condition based	Impressive bogies expertise readily available. DIANA platform making big waves for rail infra along with ADAM for stations
France	Monitoring of power and signaling of rail infrastructure (SCNF Réseau)	Condition based	After thorough testing on the most complex track, deployment was rolled out faster than originally planned
Spain	Monitoring of interlocking @ high speed lines (ADIF)	Condition based	Integral part of DaVinci infra control platform
Italy	Monitoring of rail infrastructure power distribution (RFI)	Condition based	Including augmented reality for safer and faster maintenance
Nether-lands	Real time monitoring of rail infrastructure assets (ProRail) Real time monitoring of rolling stock (Dutch Rail)	Condition based	Both nationwide and in roll out phase
Belgium	Condition monitoring of switches and power distribution	Condition based	Degradation models created by data science team

Figure 7 Current state of maintenance strategies of the major rail providers in Europe

3.4.2 Social and Ethical

Social and ethical issues that were found for this case study are discussed in the following table, which is based on Deliverable 2.3 of the LeMO Project (Debussche et al., 2018):

Table 4 Social and ethical aspects in case study 1

Aspect	Description of aspect	Relevance of issue
Surveillance	The ability to collect large amounts of data enables the data user to better understand processes in society.	Ethical implications may arise from the linking of personal data with process data. For example, in trains there is a so-called safety driving circuit which must be operated by the respective train driver at certain intervals during a journey. If the driver does not do this, the train is brought to a standstill after a previous warning tone. This prevents the driver from falling asleep. These events are recorded and can have consequences for the driver's employment. Requests for analysis of such event data in connection with personnel data are generally rejected by Siemens. However, these types of requests becoming more common.
Privacy	The involvement of privacy sensitive information	More cameras are being installed in passenger cabins. These cameras can be used, for example, to automatically recognize vacant seats. However, these systems could also be misused to monitor train personnel. The camera systems are to be used in the future for the early detection of aggressive behaviour. It is important to protect passenger identities, because simple facial recognition algorithms can be used to create motion profiles of people.
Free will	Big data analysis and associated predictive analysis that threatens human autonomy	Prescriptive maintenance gives a prescription of action. Workers or expert's opinions might be ignored and people might be forced to follow action.
Discrimination	Discrimination of population groups based on generalisations derived from big data	Due to more efficient maintenance strategy and the loss of preventive maintenance routines, maintenance workers might lose their work places.

3.4.3 Legal

Legal issues that were found for this case study are discussed in the following table, which is based on Deliverable 2.2 of the LeMO Project (Debussche et al., 2020):

Table 5 Legal aspect in case study 1

Aspect	Description of aspect	Relevance of aspect
Privacy and data protection	Compliance with privacy related legislation	Measures to comply with EU and in particular the GDPR are implemented. The collected data is mainly sensor data. Nevertheless, as mentioned above, CCTV is increasingly being implemented in passenger cabins. The latter falls inevitably within the scope of the GDPR. In any event, all data transmitted and stored is encrypted and only accessible to the corresponding data "owner".
Security	Measures to prevent security breaches	Attack and risk scenarios are played through considering attacks from outside, but also simple operational issues, such as power breakdowns for example. This is considered when building the architecture of a customer's use case to prevent most attacks from succeeding. For example, a train-to-ground connection is realised by a VPN tunnel, which is hard to hack. There are also certain mechanisms in place that make sure no man-in-the-middle attacks can be executed. Also other measures are implemented, e.g. secured access or prevention of code being pushed back into the train. Still, there are always vulnerabilities in IT systems. Therefore, a CERT (Computer Emergency Response Team) looks at systems deployed to customers and screens them for vulnerabilities and orchestrates the patching of these components. Also penetration and hacking tests are performed within Siemens Railigent. (Zicari, 2018) The security framework is aligned to the principles of industry standards, for example, IEC 62443, International Organisation for Standardization (ISO)/IEC 27001 and BSI, the German Federal Office for Information Security, and governmental recommendations for data handling in cloud environments.
Intellectual property	Claims on ownership rights of data and techniques	Algorithms for applied data analytics are considered intellectual property of the producing company and are often sold.

Sharing obligations	Compliance with legislation regarding data sharing	As a rule, Siemens concludes a non-disclosure agreement (NDA) with the respective customer. In such NDA, the sale of relevant raw data to third parties is excluded. The sale of algorithms, on the other hand, is possible and often desirable. The obligation to provide evidence applies in the event of violation of the terms of the NDA. If a violation of the NDA is suspected, it must first be proven. In any event, it will not be possible to entirely "retrieve" the data. In general, the exchange of data is always contractually regulated.
Data ownership	Disputes with regard to data ownership	Relevant data is (claimed to be) owned by the rolling stock and infrastructure operators. Access is often given only temporarily and ownership remains with the rolling stock and infrastructure operators.

3.4.4 Environmental aspect

Rail is the standard mode of transport in urban areas, at a national level, and for distances up to 1000 km. In these areas it is the most energy-efficient mode of passenger and freight transport (Mazzino et al., 2017). Rail operates with a minimal environmental impact of all modes. Carbon-free train operation and zero nitrogen oxides and particulate matter emissions are achieved by several train operators in Europe.

Using Big Data Analytics to make the rail sector more compatible to other modes like plane, might have a huge impact on the greenhouse emissions in the mobility sector in Europe.

3.5 ASPECTS RELATED TO OPPORTUNITIES, BARRIERS AND LIMITATIONS

3.5.1 Opportunities

The opportunities presented by improved predictive maintenance on the basis of big data are:

- More punctual and reliable train services.
- European-wide cooperation of involved rail operators and other stakeholders.
- Cost decrease of train services for passengers and logistic operations.
- Higher competitiveness of rail to other modes like air travel.
- Reduction of greenhouse emission in the transports sector.

3.5.2 Barriers and limitations

Nevertheless, the most important barriers are the following:

- Different infrastructures like rail and grid systems throughout Europe.
- Data sharing and handling is not standardized.
- Predictive maintenance might cause large investment costs.
- Algorithms might contain errors, but be trusted anyway, since they are black boxes.

- Expert knowledge required to implement new maintenance strategies, lack of skilled workers.

3.6 SOURCES OF INFORMATION

3.6.1 Interview

Interviews were conducted with the senior members of the Siemens AG:

- Vice-president of Data Services
- Implementation Manager
- Head of Business Development and Strategy

3.6.2 References

Arbour, G., 2019. Maintenance strategies [WWW Document]. Fiixsoftware Blog.

Bandyopadhyay, D., Sen, J., 2011. Internet of Things: Applications and Challenges in Technology and Standardization. Wirel. Pers. Commun. 58, 49–69. https://doi.org/10.1007/s11277-011-0288-5

Buurmeijer, M., Whitmore, B., Dua, N., 2018. Innovative, Intelligent, Seamless and Safe. Capgemini Ned. B.V.

Debussche, J., César, J., De Moortel, I., 2020. Report on Legal Issues, LeMO Project, D2.2. Eur. Union's Horiz. 2020 Res. Innov. Program.

Debussche, J., César, J., Hong, M., Nordeng, Z., Waldenfels, R., 2018. Report on ethical and social issues.

Hee, K., Ivanov, T., Zicari, R. V., Waldenfels, R., Özmen, H., Mushtaq, N., Hong, M., Teoh, T., Akerka, R., 2018. Big Data Methodologies, Tools and Infrastructures.

Lugara, A., 2018. Use IoT To Advance Railway Predictive Maintenance. Hitachi White Pap. 28.

Macário, R., 2011. TURBLOG: Conceitos e Modelos de Negócio para a Logística Urbana.

Mazzino, N., Perez, X., Meuser, U., Santoro, R., Brennan, M., Schlaht, J., Chéron, C., Samson, H., Dauby, L., Furio, N., Hernandez, C., 2017. Rail 2050 Vision - Rail - the Backbone of Europe's Mobility. ERRAC, Eur. Rail Res. Advis. Counc.

Sammut, C., Webb, G.I. (Eds.), 2010. Unsupervised Learning, in: Encyclopedia of Machine Learning. Springer US, Boston, MA, p. 1009. https://doi.org/10.1007/978-0-387-30164-8_867

Siemens AG, 2019a. Siemens Railigent [WWW Document]. Siemens Website.

Siemens AG, 2019b. Mobility Services. Siemens White Pap.

Siemens AG, 2018. MindSphere. Siemens White Pap.

Siemens AG, 2016. Velaro: Top performance for high-speed. Siemens White Pap.

Stern, S., Behrendt, A., Eisenschmidt, E., Reimig, S., Schirmers, L., Schwerdt, I., 2017. The Rail Sector's Changing Maintenance Game. McKinsey Study 22.

Zaldivar, I., 2017. New business models for ITS: D3.1 Market Research Analysis. Eur. Union's Horiz. 2020 Res. Innov. Program.

Zicari, R. V., 2018. On AI and Data Technology Innovation in the Rail Industry. Interview with Gerhard Kress [WWW Document]. ODBMS Ind. Watch.

4 Case study 2: "Open data and the transport sector"

The case study on "Open data and the transport sector" focused on the use of big data facilitated by open data policies and practices in the road transport sector. The practices of two organisations in Norway: Statens Vegvesen (The Norwegian Public Roads Administration) and Entur A/S (owned by The Norwegian Ministry of Transport and Communications), were studied along with other open data related initiatives in Europe.

4.1 CASE STUDY OVERVIEW

4.1.1 Organisation/Stakeholders

The supply chain of the transport business contains a series of actors, playing various roles in facilitating services associated with trade or providing a supporting facet, for instance:

- **Service-providers/owners:** Parties that own vehicles and make decisions on how to use existing vehicles to provide transportation services, when and how to buy new vehicles and what vehicles to buy.
- **Builders:** Parties that build and repair vehicles and sell them to service providers-owners.
- **Classification Societies:** Parties that verify that the vehicles are built in accordance to their own Class Rules and verifying compliance with international and/or national statutory regulations on behalf of administrations.
- **Component Manufacturers:** Parties that produce all pieces of equipment and material for the vehicle.
- **Consultancies:** Parties offering design and superintendence services to transport-owners.
- **Transport administrations/authorities:** Agencies within the governmental structure of states responsible for transportation affairs such as driver or vehicle licences, infrastructure maintenance and so on.
- **Terminal operators:** Parties that provide services such as berthing and cargo handling.
- **Charterers:** Entities that employ vehicles to transport cargoes.
- **Transport associations:** Entities providing advice, information and promoting fair business practices among its members.

Furthermore, there are numerous actors to make this industry sector functioning, such as fuel provider, crew leasing companies, transport academies, etc.

In this case study, we have focused on two organisations responsible for road transport in general and for public transport. The table below provides an overview and a rough qualitative categorisation of their business with respect to data use.

Table 6 Interviewees of the case study

Organisation	Designation	Knowledge	Position	Interest
Statens Vegvesen (Norway)	Senior engineer, Data analyst	Road transport	Acquisition Storage Usage	Road maintenance Optimal utilisation of infrastructure

Entur A/S (Norway)	CTO Data analyst	Multi-modal transport	Analysis Curation Storage Usage	Improve multi-modal mobility service Increase access to public transport data

4.1.2 Sector/Mode

This case study deals with open data as it intersects with big data in transport, in general, with special emphasis on road transport and public transport.

4.1.3 Case study motto

This case study investigates the potential for open data to support and enhance big data availability and effective application in the transport sector.

4.1.4 Executive summary

Open transport data has the potential to greatly benefit both citizens and the public sector. Public transport will only be widely used if the service is effective, efficient and user-friendly. Information that can help citizens plan their journey in advance can help enhance service and ultimately facilitate greater use. The open data movement has its conceptual roots in the international scientific community's World Data Centres which have provided open access to scientific data since the 1950's[3]. More recently, and particularly in the last 5 years, a growing amount of data has been made openly available by authorities, companies and governments around the world. Open data are available for anyone to use, for any purpose, at no cost. The degree of openness of a dataset can be rated based on the 5-star scale proposed by Tim Berners-Lee, the inventor of the Web and Linked Data initiator. The characteristics that differentiate open data from other available data are: open data are typically 'cost free' to consume and re-use. Open data are readily discoverable and supported by meta-data that make it easier for machines (e.g. search engines) and humans to find. Open datasets are free from licensing restrictions, and typically the only requirement for re-use is attribution and commitment to onward sharing. Open datasets are free from licensing restrictions, and typically the only requirement for re-use is attribution and commitment to onward sharing.

This case study will explore challenges and opportunities of open big data in transport sector with respect to various perspectives such as technological, economic, legal, policy, societal, ethical and political aspects.

[3] Trusted Data Services for Global Science , http://www.icsu-wds.org/organization

4.2 TECHNICAL PERSPECTIVE

4.2.1 Data sources/Uses

Transport related datasets can be grouped into the following themes: Place & Space, Environment, People, Things & Movement, Disruption and event-related data, Public Transport Services, Personal Automobility, Cargo connections, International Connections, and Consumption & transaction data.

There are datasets that will likely drive the rise of intelligent integrated transport: map data; weather; personal location data; network disruptions; planned events; real-time network capacity for people, vehicles & goods; public transport schedules; vehicle location data; fare and pricing data; sentiment data from service users and non-users; third party service usage data; and payment/transaction data.

Within these core datasets we recognised apparent transport-related data gaps which could be filled with varying degrees of effort. In some cases, these datasets do not exist at all (e.g. on street parking availability), in others the data exist in silos (e.g. automated cycle count data, Urban Traffic Management and Control traffic flow data) or are not open/available (historic passenger ticketing data). Most commonly the datasets are only localised, rather than consolidated at a national level (e.g. databases of major events held by local authorities). In almost all cases the technical challenges to making these data available are secondary to data owners' attitudes, costs of establishing and maintaining sensor networks, in-house skills needed to support data sharing, and data privacy concerns. The following table lists transport-related open data with the usage objectives and the 5Vs of Big Data. Note Big Data dimensions are various according to providers' business. For example, data from closed-circuit television (CCTV) could be updated per every five min or one hour.

Table 7 Main data sources of the case study

<table>
<tr><th>Data source</th><th>Used in</th><th>Big Data dimensions</th><th>Other remarks</th></tr>
<tr><td>Traffic data from sensors on roads</td><td rowspan="6">road planning & design
road maintenance & operation
environmental tasks
pollution from traffic (dust and noise)
landscape planning
traffic safety work
statistics about traffic accidents</td><td>Volume, Velocity, Veracity</td><td>Vehicle number, speed, light signal, traffic flow etc.</td></tr>
<tr><td>Weather data on roads</td><td>Volume, Velocity, Veracity</td><td>Weather, wind, temperature, snow in air and roadway.</td></tr>
<tr><td>Photos and videos from road CCTV cameras</td><td>Volume, Variety</td><td>From a part of CCTV tested for personal information.</td></tr>
<tr><td>National road data including toll, tunnel, road signals, etc.</td><td>Volume, not dynamic</td><td>National, county, municipal, private and forest roads. Physical road network.</td></tr>
<tr><td>Parking information</td><td>Not big and dynamic</td><td>Parking areas and capacities.</td></tr>
<tr><td>Traffic accidents</td><td>Dynamic, not big</td><td>Accident, vehicle, users' levels etc.</td></tr>
</table>

Schedule for public transports		Not big and dynamic	
Vehicle information		Dynamic, not big	Car types and stocks, etc.
Various reports		Not big and dynamic	Statistics data
Environmental data on road		Volume, Velocity, Veracity	Noises, emissions, etc.

As a government-owned transportation company and one of National Access Points of Norway, Entur AS (owned by The Norwegian Ministry of Transport and Communications) is currently implementing the NeTEx standard[4] with the Norwegian Railway Directorate for all national public transport in Norway. The goal of this initiative is to entirely include NeTEx for all data exchange of public transport information across Norway and collect all public transport data in one national database. Also, a NeTEx/IFOPT compliant national database for Stop Places and Points-of-Interest is established, serving as the master data source for all fixed objects including equipment and facilities used for public transport and planning thereof. In a similar fashion, real-time feeds are exchanged using SIRI interfaces and collected in a national real-time proxy.

All Norwegian public transport data will be open to the public, both as a journey planner API and as open data in Transmodel based formats (NeTEx/SIRI). In addition, Entur AS has launched a national journey planner, based on these data. This service is available to the public via website and smartphone apps. The first version of the back-end platform and journey planner was launched in November 2017, with services covering nationwide journey planning for all modes of public transport as well as basic ticketing for railway services within the greater Oslo region.

As of now, 18 of the ~25 major providers of public transport data exchange their data in NeTEx. It is expected that all major providers exchange their data in NeTEx by the end of 2019. In addition, Entur integrates 25 sources of real-time data over the SIRI interface, covering most regions in Norway.

SV plans, builds, operates and maintains national and county roads in Norway. And they comprise the Directorate of Public Roads and five regional units - Northern Region, Central Region, Western Region, Southern Region and Eastern Region – and has 72 Driver and Vehicle Licensing Offices and five Traffic Control Centres distributed across the country. SV is responsible for managing traffic information of Norwegian road and carrying out driver tests and inspection of vehicles and road users.

Dynamic data (e.g., road condition and congestion), static and historical data of road (all together 206 000 km [5]) in Norway are publicly provided by SV in near real-time. The data can be used for various purposes: to store correct data with the right quality; make better user

[4] Network Timetable Exchange - Standards Context: Transport Standard Context, http://netex-cen.eu/?page_id=58

[5] The National Road Database - Statens Vegvesen, https://www.vegvesen.no/en/professional/roads/national-road-database

interface and easier access to important data for both internal and external users; to give better possibilities for different presentations and analysis; to be used for standard report tools; to establish a new data model for the road network; to give a common feature catalogue; to use standard GIS tools on the market; especially to maintain physical road network.

On the other hand, SV has implemented DATEX II (specification for DATa EXchange between traffic and travel information centres), which is a European standard for the exchange of traffic information a well-known information model (D2logicalmodel[6]) using the format Extensible Markup Language (XML), as standard format for the exchange of real time road and traffic data. Real-time road weather data, travel times, CCTV images and incidents from the national Traffic Information Centres (TIC)-system (including closed roads and road works) are available in DATEX II format, free of charge for media and service providers. It is in accordance with Norwegian Licence for Open Government Data, NLOD[7].

4.2.2 Data flows

There are five key mechanisms for data creation (**manual collection, overt crowd-sourcing, covert crowd-sourcing, sensor-derived, service provider generated**). Web-connected fixed and mobile sensors, plus crowd-sourcing are the emerging transport data collection mechanisms. In particular, personal/vehicle location data are important for intelligent mobility.

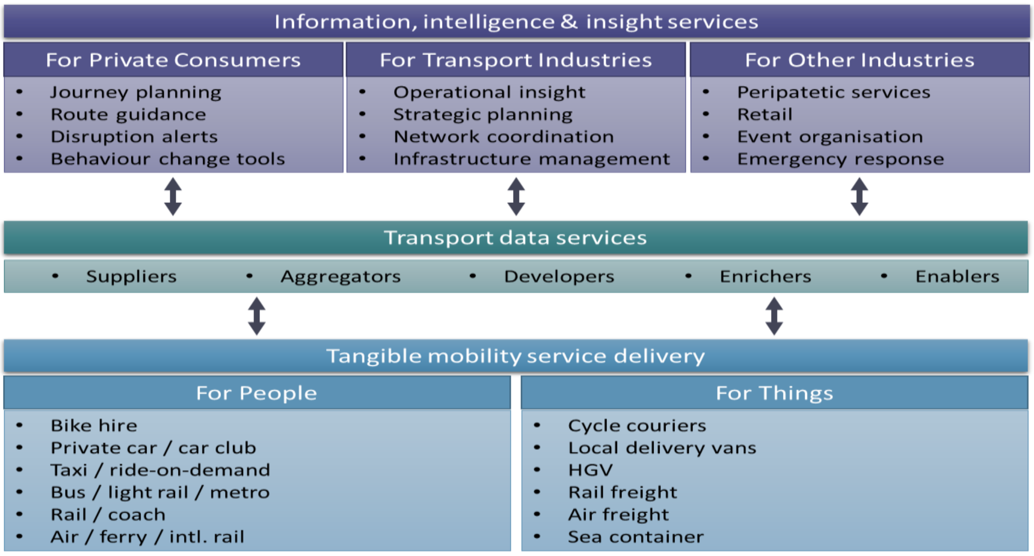

Figure 8 Data flow of big transport data

The availability of personal/vehicle location-based datasets is being driven by the private sector, which plays a growing role in collecting and aggregating transport-related data. The increasing number of partial datasets creates a 'signal problem' – because we know less about what the data being gathered represent (e.g. in terms of sampled population, demographics). Data validation and curation become very important, but few public sector data owners have

[6] D2LogicalModel, http://www.vianova-eureka.no/datex/2.3/index.htm

[7] Norwegian Licence for Open Government Data (NLOD) 2.0, https://data.norge.no/nlod/en/2.0

resources to do this well. Key players such as Norwegian Road transport authority are leading the way, while many local authorities are still developing digital strategies. Some are replicating work being done by their neighbours and might benefit from advice and guidance on maximising their investment in open data.

When discussing relevant Big Data sources with stakeholders, mobility data was identified to be the Big Data ecosystem nucleus that will bring most of the stakeholders together. The city and the citizens are concerned with mobility and are willing to share their mobility data with technology providers if their service/solution can ensure reduction of unwanted traffic or reduction of travel times.

When mobility becomes increasingly electrified, then energy data becomes part of the bigger ecosystem. Only then a cross-optimization of energy and mobility in terms of multimodal resources becomes a possibility.

4.2.3 Relevant big data policies

This case study focuses on the presence of specific policies on open data at national level, on the existing licencing norms, and the extent of national coordination.

- **Policy framework in place at national level** provides a long-term strategic vision on open data.
- **Coordination at national level** is strong and provides guidance for local/regional levels to develop their own open data initiatives.
- **Licencing norms** are in place at national level with comprehensive guidelines and assistance[8].

National open data policy framework

Many EU countries, apart from Sweden, have an open data policy in place, and since 2017, 17 member states have updated their policy frameworks. It mainly dealt with the scope of data publication such as the definition of dataset lists, along with new priority areas and/or further specifications concerning the data licencing. In 2018, 22 Member states stated to have an open data strategy in place for the next five years. In most countries, the open data strategies seem to be embedded into the broader vision of enabling an open government, and/or smart cities and - countries.

In the UK and Belgium, open data is part of the government transformation strategy[9] as well as of the country's digital strategy[10]. Moreover, the membership in the Open Government Partnership (OGP) appears to push for publication of open data in many countries (e.g., Cyprus, France, Germany, Italy or Romania). The action plans developed under the OGP framework emphasis open data as one key priority to enable the publication of high-value datasets, to

[8] Open Data Maturity in Europe, 2018, https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n4_2018.pdf

[9] UK Government Transformation Strategy 2017 to 2020, https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020

[10] UK Digital Strategy, 2017, https://www.gov.uk/government/publications/uk-digital-strategy

foster participation and transparency and focus on measures to ensure technical assistance towards data publication at public administration level. In the case of France, new OGP national action plan advocates for an expansion of the public services such as data, APIs and open labs geared toward domains of importance, e.g., transport, energy or enterprise. In 17 Member states, the national open data policy is more ambitious than the Public Services Information (PSI) policy. For example, amongst the differences, the strengthening of governance structures (e.g., Ireland or the UK), stronger incentives for the development of open data standards and methodologies (UK), and the promotion of linked open data in the public sector (the Czech Republic). In addition to these, most EU28 countries highlight the fact that the national open data policies recommend a certain set of licences. In 25 Member states, the existing open data legislation also presents the exceptions allowing public sector bodies not to release open government data. In the Netherlands, the justifications for not opening data are transparent to the broader public. The national extension of the DCAT-AP[11] standard also includes a field in which data providers are asked to state the reasons for which that data cannot be opened to the broader public. Also, 25 Member states, apart from Austria and the UK in which all data domains are considered equally important, identified priority domains for data publication. With regards to the presence of open data policies, only Norway and Switzerland have dedicated open data policies in place. In Norway, an updated digital agenda for Norway was published in 2015, which focuses strongly on data sharing features.

National guidelines on the interoperability of open data

The second indicator analyses the existence of guidelines at the national level to enable the interoperability of data as well as boost publication and reuse. In 2018, 23 EU member states were successful in enabling the development of open data initiatives at local or regional level. It can be observed both in terms of local and regional levels running their own data initiatives (be it policies or portals), and in terms of efforts done at national level to harvest local and regional portals. This number has more than doubled compared to 2017. As shown in the figure below[12], the national portal is consolidating its position as main gateway to access the open data published throughout the country.

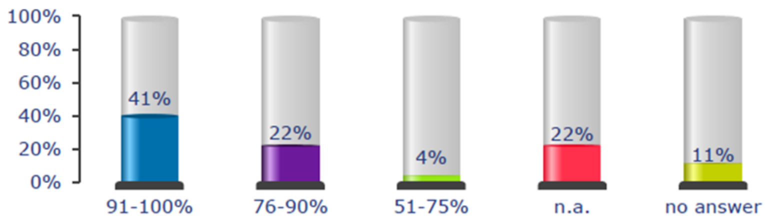

Figure 9 Local and regional portals harvested by national portal for EU28, 2018

For example, in 2018 17 countries (63%) stated that they are actively consolidating data from local and regional portals (from 76% to 100%) into their national portal. Enhancing visibility and discoverability of, and subsequently access to datasets released by the various administrations

[11] The DCAT Application profile for data portals in Europe (DCAT-AP) is a specification based on W3C's Data Catalogue vocabulary (DCAT) for describing public sector datasets in Europe.

[12] Open Data Maturity in Europe - Report 2018 by EUROPEAN Data Portal, https://www.europeandataportal.eu/en/news/open-data-maturity-europe-2018

in a country can be achieved by integrating local and/or regional portals into national portals, and hyper-local portals in cities are important players in the open data publication chain.

In order to do so, the national level needs to strike a balance between the need for a hands-on coordination and supervision to ensure national targets are met, and the need for a good level of freedom at local and regional levels. In countries with a federal structure, such coordination is constrained by the autonomy the law gives to the states. In such countries (e.g., Belgium, Germany, Denmark, Finland, Lithuania or Portugal), the national level takes a more passive approach to coordinating efforts, while most member states preferred active coordination. Such policy directions ensure to monitor the timely publication of data on the portal according a roadmap defined in the national open data programme and keep that the data quality is in line with the standards and guidelines of data formats defined in the national open data programme. That is, it's effort to push forward at a higher speed the standardisation and interoperability of data as well as publication of data on one single platform. Guidelines are also available that support local and regional public administrations with their publication processes. In 2018, only Norway and Switzerland had guidelines for data publication developed at national level. In Switzerland, the guidelines for opening data are available on the national portal. In Norway, the guidelines for data publication cover 15 areas such as zero charge, machine-readable formats, update the data, feedback channels for re-users, publishing high-quality metadata, licensing the open data and so on.

In terms of transport sector in Norway, for example, there are obligations related to national access point. It related to ITS directive[13] including actions from A to E. Action C is about safety related (meta) data. Real-time traffic information in Action B[14] is not perfectly fulfilled by Statens Vegvesen (SV). These actions are assigned to different parts of government departments. There is a portal for public data (NLOD) including from various data sources to data access from DIFI. Some of these data are more tailored to do these actions (A and B) and access points from EU. It relates to national regulations. Other actions are related to safety related information (C), harmonised provision for an interoperable EU-wide eCall (D), and safe and secure parking (E).

Licensing norms

Licensing norms refer to the extent to which guidelines are in place to ensure that the published data complies with the definition of "open" in both terms of the data being free and it being released under an open licence. Open licence is important for providing clear information on the terms and conditions under which the data can be reused, alleviates the uncertainty from the re-users' side and fosters immediate reuse. With the transposition of the PSI Directive of 2003 (2003/98/EC) into national laws, many different licensing norms were developed at national level. Creative Commons (CC) licences[15] are recommended in order to align practices across the EU and make them more transparent and predictable for potential reuses. 15 Member states stated in 2018 that all their published open data can be accessed free of charge,

---

[13] Action Plan and Directive for ITS, https://ec.europa.eu/transport/themes/its/road/action_plan_en

[14] Action B (Delegated regulation EU 962/2015) - Real-Time Traffic Information, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32015R0962

[15] Creative Commons - Attribution 4.0 International, https://creativecommons.org/licenses/by/4.0/

and another 8 EU countries provide 90-99% of data can be accessed free of charge. In terms of recommended licences, while some member states developed their own licensing terms (Germany, France, Romania or the UK), in 2018, 21 EU member states promoted the use of CC licences. Apart from Liechtenstein, all European Free Trade Association (EFTA) countries stated that all their published data is published free of charge. Additionally, in the three countries (Iceland, Norway and Switzerland), guidelines were published at national level to help public bodies correctly licence the published data.

4.2.4 Main technological challenges and issues

As realised in the interview with SV, data quality is one of the most important challenges for open data in the transport sector. Different transport operators in the field of data generation to curation, face critical challenges such as data incompatibility and needs for ensuring data quality. As a result, good architecture for the integration of external data is most needed.

In addition to these, standardisation and interoperability of data is critical issue. The DCAT-AP standard focuses on licensing, copyright and law restrictions. The preference for a more active approach is showcased by the Member States' effort to accelerate the standardisation and interoperability of data as well as the publication of data on one single platform.

There are other issues such as data flow control related to roles and data ownership, making the collection and publication of data more complex and difficult and making it more difficult to ensure the quality of open data.

4.2.5 Big data dimensions and assessment

In terms of velocity and veracity of Big Data, quality dimensions of open data can be measured by a **systematic automation and data & metadata currency** for harvesting and the accuracy & reliability of available data.

The automation measurement refers to the availability and effectiveness of the approach used by countries to ensure that the metadata describing the available datasets is updated regularly. In 2018, 20 of 26 national portals state that a predefined approach is in place to ensure the currency of metadata and data. Some data in different levels is produced at a precise frequency and harvested automatically. For example, in 2018, only Italy and Belgium stated that all metadata is uploaded in an automated way to the national portal. In another 5 countries (the Czech Republic, Germany, the Netherlands, Spain and Sweden) over 90% of metadata upload takes place automatically. In terms of machine re-usability[16] of datasets, 17 of 26 portals stated in 2018 that over 90% of the data discoverable on their national portals is available in a machine-readable format. Another 8 national portals stated that between 71 and 90% of the available datasets are published in such formats (RDF, XML, JSON and so on). In 2018, the number of countries that have provided 70% or less of their data in machine-readable formats

[16] In link with the definition provided in the revised PSI Directive (Directive 2013/37/EU), 'machine-readable format' refers to a file format structured so that software applications can easily identify, recognise and extract specific data, including individual statements of fact, and their internal structure.

dropped. It can be improved by the training activities that were organised at national level to help data publishers improve their data publication process as well as the bilateral support that many countries are supporting to national public administrations to publish higher-quality data. In terms of the automatic updating of the (meta-) data on the national portal, Iceland and Norway stated that this occurs in over 90% of cases. Switzerland has lower automation and it ranges between 10 and 30%.

The currency of data is often critical to application. Analysis and insight are often not relevant when performed on outdated data used to describe a phenomenon that requires tackling today. The frequency of updates depends mainly on the type of data being published. Whereas timetable data of public transport, for example, needs to be updated only on the occurrence of new changes itself, other types of data such as traffic flow, accident and weather data need much more frequent updating. Applications such as intelligent journey planners, for instance, simply cannot be built without up to date, if not real-time data. Instead metadata currency is critical to reuse and interoperability. A dataset that is described by old metadata - that, for example specifies incorrect names or wrong types for a table's fields - will most likely break any sort of automation that is built to use that data, and hinder both the dataset and the catalogue's discoverability. For metadata currency assessment, portals might be asked to indicate the update frequency of the metadata describing the datasets offered through the national portals. Questions follow the DCAT-AP guidelines with regards to the field "update frequency" and categorise currency against the following options: less frequently than monthly, monthly, weekly and daily. In terms of currency of the data and the availability of historical versions, there appears to be a good balance. For quality of data, Norway and Switzerland recorded a machine-readability of over 90% of the featured datasets. In Iceland, the level of machine readability ranges between 71 and 90%.

Volume and variety of Big Data related to open data seems not to be carefully consider yet. For open data portal at national level, although frequently updating is one of essential issues as mentioned above, it is actually difficult to update or publish open data into national open portal without a strategic automation approach in real-time. Furthermore, even if there are automation approaches of updating open data, it is essential to accompany publishing and updating open data of local or regional portals in real-time. But in transport industry, although real companies or operators are collecting and analysing their data related to transport in (near) real-time and requiring that data and results timely, we cannot sure whether they prefer to publish their data as open data. As we discovered from the interview with SV, requirements about role ownerships of people and government have been split, meaning it is very difficult to uphold data quality for the collection and publication of data. In addition to these, data ownership could make situations between relevant stakeholders more complex.

4.3 BUSINESS MODEL

Figure 10 below shows a general model of businesses related to open data in the form of business model canvas[17]. Businesses associated to open data could collaborate or compete with

[17] TURBLOG. (2011). FP7 TURBLOG Deliverable 2 Business Concepts and Models for urban logistics.

various partners from public and private sectors, and at national, local and regional levels. Their activities involve data collection, validation and aggregation via relevant resources, techniques, policy and legal supporting in order to delivery data services and products utilising open data for clients and society. The clients range from end users to relevant business actors, and include citizens and relevant public sectors. Therefore, if other companies are customers, they can have two relationships: collaboration based on relevant directive, policy, initiative and contract, and competitive relation via license or data ownership. In the case of end users, most of relations is complementary, since they provide new data services (and products) and obtain relevant data through various communication techniques such as email, SMS, SNS, call and apps in order to create revenues and increase the number of jobs involved in producing the service or product and save costs. These benefits lead to increase utilisation of their goods and services, enhance customers' loyalty and grow potential market.

Key partners	Key activities	Value proposition	Customer relationships	Customer segments
Who are the firm's partners? Public sectors at national, local, regional levels. Private sectors from data supplier to data curators.	**What are the key activities, such that the firm operates successfully?** Data collection, validation, aggregation and collaboration with partners	**What value does the firm deliver to the customer?** Data services and products based on open data	**What type of relationship is established between the firm and the customer?** Collaboration by relevant directive, policy, initiative and contract Competitive relationship through license or ownership	**For whom is the firm creating value?** Various targets such as clients, collaborators and society
	Key resources	**Internalization of costs and benefits**	**Channels**	
	What does the firm need to create value? Proper sources, techniques, policy and legal supporting for collecting, storing and analyzing data	**How are costs and benefits from firm's business activities internalized?** Revenues and the number of jobs involved in producing a service or product and cost savings, in directly. New goods and services, time savings for users of applications using open data, efficient public services and relevant market growth, indirectly.	**How does the firm reach the customer?** Various communications techniques (email, SMS, SNS, call and mobile apps)	
Cost structure			**Revenue streams**	
What are the costs associated to the business model? It is difficult to quantify, since the cost depends on required equipment, techniques and human power.			**For what value are the customers paying and how are they charged?** Mainly increasing utilisation of new data goods and services, enhancing loyalty, growing potential market, and so on.	

Figure 10 General model of businesses related to open data

As a concrete example, the business model of SV and Entur is described as follows:

- *Value proposition*. These two government institutions collect various traffic data mentioned in Table 7 and publish as open data in order to increase access to public transport data; improve multi-modal mobility services; reduce coast for maintenance; and optimize infrastructure utilisation.
- *Customer segment*. Range of target customers is very broad from relevant transport providers, leveraging open data provided, to citizens using relevant services such as travel planner.
- *Customer relationships*. As they have the broad customer segment, there are different levels of relationships. Although they have many number (approximate 7 million per year) of interaction with end-users (i.e., citizen), feedback being in fact used in their business process are relatively small. Impactful feedback are from relevant business partners (e.g., other authorities, transport providers, etc.)

- *Channel.* Many methods to communicate with customers such as email, SMS, call, chatting and apps[18]. For example, service providers or developers interact through the emails mentioned their app[19] and API documents[20], while end-users who inform a road accident could contact by calls. However, most of the calls from the users are related to current traffic conditions, thus it might make capacities of their business exhausted rather than improve business processes.
- *Key activities.* Key activities to make the value proposition are collecting traffic data and data associated with transport; aggregating the data; validating the data to improve data quality and to check legal aspects such as privacy and security; publishing the data as open data on their[21] or national portals[22]; In addition, to generate and publish more useful information, they collaborate with other companies (e.g., NSB and Kringom), counties (Sogn og Fjordane) and regional communities.
- *Key resources.* Although the infrastructure for collecting and transferring data is established on road, there are huge needs of human resources to maintain the infrastructure and data quality gathered. Therefore, it depends on budget funded from government. However, there is relatively no difficulty for gathering and publishing open data, since SV and Entur have already their standard (DaTex II and NeTEx) according to EU standard.
- *Key partners.* Key partners are also vary, since the range of target data related to transport are large. In the case of Entur considering multimodal transport provide various transportations such as rail, road, bus and ferry. SV is cooperating with other authorities maintaining road conditions and measuring weather on road.
- *Cost structure.* All costs are coming from government to generate and manage open data and to do their own tasks. However, Entur sometimes earn additional profits by providing premium services to cover the cost of human resources needed to interact with citizen.
- *Revenue streams.* Even though no direct compensation is paid for their open data, providing open data has indirect benefits such as improving traffic management, generating new product using these data and growing potential market.
- *Internalisation of costs and benefits.* As NAPs of Norwegian transport, all external costs are funded by Norway government, and the benefits of publishing open data is generating added value such as new goods & services, time savings for citizen.

4.3.1 ANALYSIS OF Business processes

Open data appears to be a valuable material but its benefits are realised through various business processes or activities such as **transformation, analysis, aggregation and synthesis**. As shown in Figure 11, value chain of open data follows business processes and includes: data creation, data validation, data aggregation, data analysis, data services and products and finally

[18] Entur – Journey Planner, https://play.google.com/store/apps/details?id=no.entur&hl=en
[19] Entur AS app info. of website, https://www.entur.org/
[20] Statens vegvesen – open data APIs, https://www.vegvesen.no/om+statens+vegvesen/om+organisasjonen/apne-data
[21] Statens vegvesen – Dataportalen-beta, https://dataut.vegvesen.no/
[22] Difi – Open public data in Norway, https://data.norge.no/

these can be further aggregated to lead to new services and added value. Following the value chain, **data is transformed into information, then into knowledge, from which services and added value can be derived**. Many stakeholders from both public and private sector play roles, such as **suppliers, aggregators, developers, enrichers and enablers**, are involved in these different phases.

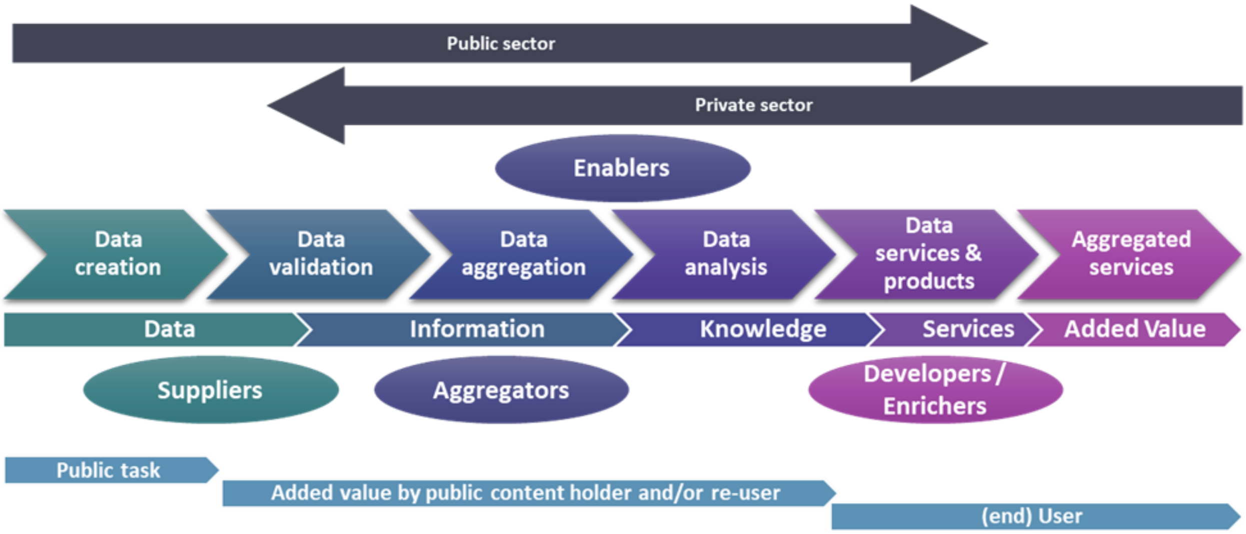

Figure 11 Business process of open data value chain

By making more data freely available for re-use, both public bodies and the private sector can create new data services and products, as well as aggregated services further downstream. To create new services and products is just one of many potential economic values. There is also the possibility of improving business processes and intelligence, creating opportunities to better match supply and demand and enabling collaborative opportunities.

4.3.2 Relation to big data initiatives

Open data is one of important issues in transport sector. There are many open transport data initiatives conducted by countries, counties and even global companies. There are some similarities in terms of dealing with similar purposes (e.g., re-use) of open data at national level. On regional or industrial levels, there are large differences according to their own business domains. This section introduces some of Big Data initiatives related to open transport data obtained from our case study and from literature.

To see the value of open data at a micro economic level, it is important to first understand what the value of open data is at the macroeconomic level. The European Union Open Data Portal (EU ODP[23]) provides access to an expanding range of data from the EU institutions and other EU bodies. The EU ODP is designed to make the EU institutions and other bodies more open and accountable in 2012. It aims to help to innovative use and unlock their economic potential.

[23] EU Open Data Portal - Access to European Union open data, https://data.europa.eu/euodp/en/home

Staten Vegvesen (SV, Norwegian: Norwegian Public Roads Administration (NPRA)) collects large amounts of data on roads and traffic per various time intervals (from real-time to near real-time). Around 40 types of datasets (e.g., traffic, weather, vehicle, environmental and parking data) related to roads in Norway is public in the formats of XML, JSON and CSV for suitable reuse (Dataportalen beta[24]). Some datasets including sensitive information could only be accessed by contacting SV. In addition, these datasets are managed by Norwegian License for Open Government Data (NLOD) created by The Ministry of Government Administration, Reform and Church Affairs (Difi[25]).

Most of authoritative UK open data including around 1,000 transport datasets are published in data portal[26] built by the Government Digital Service under the Open Government Licence v3.0[27]. In particular, Department for Transport has published 160 transport datasets in the open data portal, under Open Government Licence. The Office of Rail and Road (ORR) is operating The National Rail Trends Portal[28] as an online tool that provides access to railway statistics database. Data format vary from CSV to XML. Especially, Open Data Institute (ODI) is collaborating with companies and governments to build an open and trustworthy data ecosystem. Recently, the ODI have launched three collaborative data innovation projects between the UK and France. Etalab in France develops and runs an open platform[29] accompanying public data of the State and administrations, and around 2200 datasets have been published in the portal from 120 organisations in transportation sector. Especially, SNCF, as one of the world's biggest groups with a presence in 120 countries, opened 280 transport datasets. Formats are also various such as JSON, SHP, ZIP and CSV.

According to the Global Report[30] 2017 of the Open Data Barometer[31] produced by World Wide Web Foundation, most countries are not meeting the basic open data charter principles apart from some leading governments such as the UK, France, USA, Korea and Mexico. *Although there are many initiatives actively supported by various succour levels (such as government, municipality and county), in most cases, the right policies are not in place, nor is the breadth and quality of the datasets released sufficient yet*. Regarding these, this case study discusses challenges and opportunities, which stakeholders have really encountered with, in transport sector in various aspects of open data through literature reviews and interviews.

---

[24] Dataportalen - beta, https://dataut.vegvesen.no/

[25] Open public data in Norway - Difi, https://data.norge.no/

[26] Data.gov.uk – UK open data potal, https://data.gov.uk/

[27] UK Open Government Licence for public sector information, https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

[28] ORR office of rail and road, http://dataportal.orr.gov.uk/

[29] Data.gouv.fr – FR open data potal, https://www.data.gouv.fr/

[30] Global Report 4th Ed., https://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf

[31] Open Data Barometer, http://devodb.wpengine.com/

4.3.3 Illustrative user stories

The utilization of Big Data driven applications in transport varies largely between different transportation actors. The following stories are representative illustrations about generally encountered opinions:

- **Transport service-providers/owners**. The main interest of owners is to provide competitive transportation offerings. Data they need for their daily operations is therefore on an aggregated level, i.e. fuel consumption, emission reporting, arrival scheduling etc. The crew is responsible for smaller maintenance tasks, optimization and handling of the vehicles. The needs for data handling therefore cannot fall under the 5Vs of Big Data (Reference: D1.1); moreover, most of data collection is performed manually into forms. Few companies have established pilot cases and then only on a small fraction of their fleet to investigate automatic collection of aggregated data.
- **Transport Associations**. They do not have Big Data, nor are they keen on obtaining it. Their focus is on information sharing (on an aggregated level) and on contractual and commercial arrangements. Their concern is also on potential misuse of that data through industrial espionage.
- **Administrations/Authorities**. Transport authorities are mainly interested in aggregated data from vehicles such as environmental pollution, information about cargo, and data that is used by customs authorities etc. For example, in shipping industry these data originate from the vessels noon reporting of manual reporting when entering national waters.

4.4 SOCIETAL ASPECTS

4.4.1 Economic and political

In general, economic benefits related to open data might be data available for commercial re-use at a zero or low cost. As emphasised by the European commission in Digital Single Market Mid-Term Review, the data economy is estimated to reach EUR 739 billion by 2020. Against this backdrop, it is important to quantify the economy impact of open data in Europe. According to 'creating value through open data[32]' of the European Data Portal study, economic benefits of open data could be quantified by four indicators: direct market size, number of jobs created, cost savings and efficiency gains[33]. Open data managed by government involves two kinds of benefits:

- As **direct benefits**, there could be monetised benefits such as revenues and the number of jobs involved in producing a service or product and cost savings.

[32] European Data Portal (2015), Creating Value through Open Data, Study on the Impact of Re-use of Public Data Resources, https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

[33] European Data Portal (2017), Re-Using Open Data, A Study on Companies Transforming Open Data into Economic and Societal Value, January 2017. https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf

- **Indirect economic benefits** relate to new goods and services, time savings for users of applications using open data, efficient public services and relevant market growth.

In this regard, open data could offer a substantial economic and societal potential. However, this can only be realised when the data is accessible, re-usable and re-used by the stakeholders creating value from it. According to report analysing re-usage of open data, transport sector has shown noticeable correlations with other fields such as population & society (68% chance), regions & cities (60% chance) and environment (52% chance). In addition, transport open data has been identified by the European Commission as one of the thematic data domains that are expected to generate the highest demand from re-users across the EU[34].

Political impacts of open data are considered as transparency and efficiency of government and public service delivery, and so on. Most EU member states have shown positive responses (around 55% political impact) about impact of open data for these transparency and efficiency in 2018[35]. In particular, from several cases such as AppValencia[36] in Spain, Entur – Reiseplanlegger[37] in Norway and so on, the transport sector seems to be one noticeable field for improving government efficiency and encouraging public service.

4.4.2 Social and Ethical

Open data could contribute to society in general and on the inclusion of marginalised groups in particular. For example, in transport sector, usability information (e.g., elevators or escalators to bypass hindrances for destinations, public transport vehicles including spaces for wheelchairs, etc.) could be critical for people with disabilities or minority groups. For example, some projects are the social inclusion and community activation programme[38], the ability programme[39] or gender equality[40]. There are also plans for the publication of building information management data as open data. In terms of impact for society in general, as other example, France is committed to "Data for Good" initiatives involving data scientists, developers and volunteers who offer their digital skills to serve social projects[41]. In this context, a tool to highlight impact of open data on the social dimension was developed and presented early 2018, on the occasion of the Open Data Day[42]. Regarding the extent to which the civil

34 European Commission (2014), guidelines on recommended standard licences datasets and charging for the reuse of documents. https://ec.europa.eu/digital-single-market/en/news/commission-notice-guidelines-recommended-standard-licences-datasets-and-charging-re-use

35 Open Data Maturity in Europe - Report 2018 by EUROPEAN Data Portal, https://www.europeandataportal.eu/en/news/open-data-maturity-europe-2018

36 Valencia - La NUEVA APP,https://www.valencia.es/ayuntamiento/atencion_ciudadano.nsf/vDocumentosTituloAux/Aplicaciones%20m%C3%B3viles?opendocument&lang=1

37 ENTUR - Entur-appen, https://www.entur.org/entur-appen/

38 Social Inclusion and Community Activation Programme (SICAP) 2018 – 2022, https://www.pobal.ie/programmes/social-inclusion-and-community-activation-programme-sicap-2018-2022/

39 Ability Programme, https://www.pobal.ie/programmes/ability-programme/

40 Gender Equality, http://eufunding.justice.ie/en/eufunding/pages/genderequality

41 Data For Good, https://dataforgood.fr/

42 Data-story #1 Handisco (Nancy), https://www.data.gouv.fr/fr/posts/data-story-1-handisco-nancy/

society projects are supported on the social dimension, half of EU member states are supporting such initiatives in 2018. *Most of social impacts of publishing and generating open data look optimistic.*

On the other hand, there are some negative concerns in terms of ethical impact of open data. Ethical issues raised by the personal data which could be included in open data are complex, especially in transport sector, and as a result some organisations are exploring the rapidly emerging discipline of data ethics based on the ODI. For example, some data such as passengers' ticketing, waiting time, CCTV or vehicle registration has very sensitive information when they are published as open data. In our case study with SV, there was concern about many cameras on roads. But it is under very strict management to keep anonymous, is not used for identifying private properties and is not public. After testing cameras in terms of the anonymous, some of cameras are public, but there are still many un-public cameras. In other case, there might be side-effect of open data, for instance, if we open data related to people driving above speed limit, it could encourage more people drive over the speed limit. Therefore, it could help to know usage purpose for making appropriate open data. In addition to these, in terms of open data, misuse must be carefully considered, since it might happen serious problem, for example, malicious usage of information related to traffic light.

The table below lists the societal and ethical issues in businesses of SV and Entur according to the six topics identified in D2.3 “Social and ethical issues”.

Table 8 Societal and ethical issues in businesses of SV and Entur

Aspect	Description of aspect	Relevance of issue to SV and Entur
Trust	Increase of decrease in trust in by or between the data users, data facilitators and data suppliers.	The collaboration between relevant organisations might increase data trust by aggregating more and more traffic data. Although various formats of data make data quality lower and then it links to decrease of data trust, SV and Entur have effort to give uniform standard to solve the problem. And it is progressing slowly and steadily.
Surveillance	The ability to collect large amounts of data enables the data user to better understand processes in society.	Both of authorities wanted to know flows of open data to avoid misuse of the data, but they think that it is impossible. Instead, they are thoroughly confirming the data as enough as possible before making them public.
Privacy	The involvement of privacy sensitive information	In the case of SV, they collect photos and videos from CCTV on roads and publish the data to create added value. Since these data could have personal information, this kind of data is carefully checked whether the data includes personal properties and information.

Free will	Big data analysis and associated predictive analysis that threatens human autonomy	Providing open data does not need sophisticated data analysis, therefore these might be no threatens of free will. However, there is need of a method to track usage flows of open data, since the possibility of misusing open data always exist.
Personal data ownership	Confusion of who owns the data or loss of data ownership by data suppliers	Due to agreements between the transport sector and government about the usage of relevant traffic data, the ownership of this data is not considered a big issue. However, when the collaboration expends with commercial companies, data ownership should be considered to avoid any confusion of owning original data.
Discrimination	Discrimination of population groups based on generalisations derived from big data	No issues related with discrimination were found.

4.4.3 Legal

The EU institutions have taken both legislative and non-legislative measures to encourage the adoption of open data, most notably through the PSI Directive which attempts to remove barriers to the re-use of public sector information. A proposal for a revision of the PSI Directive intends to extend the scope of application to public undertakings, including actors in the transport sector.

As aforementioned, it is relatively easy to contain personal information into open data. With regards to **privacy and data protection**, some concepts, principles and obligations under data protection law appear to be problematic for the uptake of Big Data, especially open data. In particular, the broad definition of "personal data" and "processing", the qualification of the various actors involved as (joint-)controllers or processors, the core data protection principles, the need to identify a ground for processing, the requirement to conduct data protection impact assessments, the implementation of privacy by design and by default measures, the rights of data subjects, and the requirement to put in place adequate data transfer mechanisms seem difficult to reconcile with the concept of Big Data. As case study with SV, we confirmed again that privacy is one of the critical issues for open data. For example, one possible problem is combining several datasets that might be used to identify certain people. There are many initiatives to extract personal information from public data. Once data is published, it is difficult to track the usage of the data.

In this context of **sharing obligations**, while private companies often generate huge amounts of data, they are not always prepared to voluntarily share this data outside the company. This is due to the large number of legal, commercial and technical challenges associated with private sector data sharing. On the other hand, it seems to encourage data sharing at the national level and some EU member states (e.g., Norway and Switzerland) are following relevant EU directive. For example, there is one EU initiative to make national access point for encouraging open data. Implementing of that EU directive is different for countries, some of them have already released with high or enough quality levels. SV has intermediate solution by the directive of information Directorate for Administration and ICT[43]. The legal issues identified according to the framework of D2.2 "Legal issues" is shown in the following table.

Table 9 Legal issues in businesses of SV and Entur

Aspect	Description of aspect	Relevance of aspect to SV and Entur
Privacy and data protection	Consideration and implementation of the requirements of the GDPR and other types of privacy and data protection-related legislation	Both authorities made significant efforts to comply with privacy legislation and are waiting the enactment of local legislation.
(Cyber-)Security	Awareness and implementation of certain general or specific (cyber-)security requirements	They have a special team whose task it is to comply with rules and laws with regard to security and cyber-security. In the case of Entur, cloud servers, including mature security functions, are being used to store traffic data.
Anonymization and pseudonymisation	The use of any type of anonymisation or pseudonymisation technique, i.e. techniques to ensure personal data can no longer be attributed to individuals without using additional information or to reduce the identifiability of individuals	Although they in principle do not process any personal data, data which could identify a data subject are anonymised.
Supply of digital content and services	Uncertainties regarding supply of digital services for personal data	Since there is no relevant regulation forcing them to publish open data within a certain period of time or moment, using such data might be difficult to actively use for a relevant service.
Free flow of data	Awareness of any data localisation requirements hindering the free flow of non-personal data for the organisation	No data localisation requirements are applicable to both authorities, however some open data are still published in Norwegian.

[43] Directorate for Administration and ICT (Difi), https://www.difi.no/

Intellectual property	Protection by the organisation of its service/product through intellectual property rights and hindrance for the organisation of third parties claiming or otherwise asserting intellectual property rights	The data are available as open data. With public organisations involved, future claims of intellectual property are unlikely.
Open data	Awareness and use of open public sector information	Since both authorities' open data are not considered to include sensitive information, its availability is not at risk.
Sharing obligations	Awareness of and compliance with certain specific data sharing obligations, mandatory legal requirement to share data with other market actors	As NAP of Norway transport, both authorities comply with the relevant EU and Norwegian legislations.
Data ownership	Third parties claiming data ownership	Since public organisations are involved, data ownership claims might be unlikely. However, it is possible some issues are caused once the generation of open data is more common sometime in the future. Therefore, relevant regulations should still be considered.
Liability	Use of data from non-public organisations and the organisation of the data sharing	The consequences of wrongly displayed information can be significant. Correct information is important to ensure data quality and trust. Therefore, both authorities uphold government supported standards and are making efforts to try ensuring data quality.
Competition	Consideration of legal liability framework and any particular issues in this respect	No issues related to competition are part of their businesses.

4.4.4 Environmental

In transport sector, from our case study with SV, we realised that there are many initiatives related environmental and transport sectors. *In terms of environment aspects in transport sector, there might be emission of gases, particles and noises from roads and traffics.* It could be reduced when taking right problems as a big impact of Big Data. Regarding GeoFencing, SV is trying to reduce the emissions based on traditional modelling statistics, although it might be not in real-time and optimal way. **GeoFence EU project** as zero emission zone, collaborates with Volvo cars, demonstrate in Oslo that showed how we can use geofence, or a virtual fence in which hybrid cars switch on petrol to electricity, to reduce emissions in 2018[44]. It could be used for emissions as well as speed. *On the other hand, in real world, there might be some negative issues, related to these project, from citizens or politicians. For instance, some people could be furious when they are restricted for their freedom about speed on roads.* Some pilot projects are combining data sources from more sources to deal with cars having bad status of equipment in winter. It could be a problem related to personal privacy issue, but is also linked to safety and dangerous vehicles on roads. In this context, such projects like GeoFencing should consider the social impact.

In terms of environmental impact of open data, we couldn't find severe negative concerns or side effects, instead 12 EU member states stated that they supported civil society initiatives that aimed to identify policy solutions to environmental challenges faced by the country since 2017. For example, SV are collaborating with Meteorologisk Instututt to obtain weather information (e.g., weather data processed or forecasted) by measurement stations. Since the stations are distributed on large areas, some of them is owned by SV and municipalities. And it is provided by API. In addition, Norwegian Institute for Air Research[45] was collecting air quality in past, now they are responsible for different roles with new project. Outcome from analysing air condition and historical data are used by typical consultants related to air quality considerations[46]. Based on the experience, it is planning to make an impact calculator for assessment of projects on air quality and usage as official tools. This kind of initiatives could allow to provider better information and recommendations to policy makers for environment.

4.5 ASPECTS RELATED TO OPPORTUNITIES, BARRIERS AND LIMITATIONS

4.5.1 Opportunities

The range of open data is expanding, improving in quality, and becoming more dynamic through live feeds that are constantly updated. Even though the increased range and scope of the datasets becoming available makes their coordination more challenging and can even make it harder to discover and exploit high-value transport data feeds, with supporting from many initiative for policy and legal aspects at the national level, open data could give society valuable opportunities such as **creating and extending new data services, products and markets**.

This is noteworthy because in the future, all transport companies are also expected to be data businesses – exploiting the '**digital exhaust**' from their operations and customer interactions. Many of the major global data and technology companies are already investing in transport systems to explore whether they can provide enhanced mobility services. These will likely draw on the **deep insights** their analysis of customer data reveal on individual-level and aggregated travel intentions, actions, and purposes. The dominance of major technology companies and vehicle manufacturers making open data, that will make more open data from their own data, means they are best-placed to crowd-source and exploit large datasets, and drive the

[44] Geofence as zero emission zone - Statens Vegvesen, https://www.vegvesen.no/fag/trafikk/its/geofence-til-nullutslippssone-demo

[45] Norwegian Institute for Air Research (NILU), https://www.nilu.no/

[46] Air quality in Norway, https://luftkvalitet.miljostatus.no

globalisation of standard data formats that will support more user-focused and integrated transport systems.

4.5.2 Barriers and limitations

Standardisation of data remains an issue, with few public and private sector providers currently adhering to globally common data formats, and some forms of real-time data standards still to be defined. Developers and service providers are increasingly interested in globally common data formats (e.g. GTFS, DATEX II in Europe), which make it easier to scale products and services worldwide, yet few private sector providers adhere to them. Some forms of real-time data standard are yet to be defined (e.g. personal location-based data feeds).

More must be done to address public concerns over **personal data privacy** – identified as one of the key long-term threats to creating successful crowd-sourced products and services. For personal data, a balance must be struck between the benefits of providing access to data – and individuals' rights to privacy and being protected from harm. It means this data can only be shared in safe ways (encryption, anonymization, etc.) with user agreement and, if the person has not consented, not at all.

Transport data discovery and its uneven **availability** is becoming a major challenge. The multitude of catalogues and platforms needs to be embraced and documented. The best transport data feeds are concentrated at strategic national/major city levels. Furthermore, almost no private sector data is documented in any form of catalogue. There are very little real-time data feeds are openly available.

Table 10 Opportunities and challenges in businesses of SV and Entur

Opportunities and enablers	Barriers and limitations
Creating and extending new data services, products and markets. For example, open data generated by SV and Entur promotes service providers to develop services such as the travel planner considering conditions of road and public transport. *Big Data technology.* Whereas the volume and variety is relatively small, the velocity of poses more work, with veracity being the most challenging component. Based on open data as a start point, both authorities are actively collaborating with other organisations with experience and resources to handle such data technologies.	Personal data. Both authorities have many effort to deal with privacy issues. When range of open data is expanded, possibility of including personal data into open data will be increase. Thus, this issue should be continuously considered, and relevant techniques should be developed to find appropriate balance between benefit of open data and protection of personal data.

Exhausting digital value and getting deep insight in transport sector. Standards used by SV and Entur promote to generate more digital value for public sides. It will lead more deep insight when their open data meets with service providers who have capacity of using Big Data technologies to obtain insight (e.g., predicting traffic congestion and foreseeing accident) in transport sector.	*Extending standard.* Although they already have standard and follow EU standard based on relevant regulations (e.g., ITS action plan), there will be needs for more standard with increasing real-time data and developing new technologies such as autonomous vehicles. It means that proper standard should be adapted for future technologies. *Localisation and internationalisation.* Standardisation for open data also need to have ability adjusting to local environment. For example, Norway has many mountains and valleys, it has led to development of air and ferry transport. As a result, useful open data could be different for each region and country.

4.6 SOURCES OF INFORMATION

4.6.1 Interview

Two interviews with Statens Vegvesen were carried out with

- Senior Engineer
- Data Analyst

Two interviews with Entur A/S were carried out with:

- Chief Technology Officer
- Data analyst

4.6.2 References

4thEdition-GlobalReport – Open Data Barometer (ODB), 2017, https://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf

Ability Programme, https://www.pobal.ie/programmes/ability-programme/

Air quality in Norway, https://luftkvalitet.miljostatus.no

Analytical Report 10 Open Data and Entrepreneurship, 2018, https://www.europeandataportal.eu/sites/default/files/analytical_report_10_open_data_and_entrepreneurship.pdf

Analytical Report n9 The economic benefits of open data, 2017, https://www.europeandataportal.eu/sites/default/files/analytical_report_n9_economic_benefits_of_open_data.pdf

Big and Open Data in Transport - Parliament UK, 2014, https://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-472

Creative Commons - Attribution 4.0 International, https://creativecommons.org/licenses/by/4.0/

D2LogicalModel, http://www.vianova-eureka.no/datex/2.3/index.htm

Data For Good, https://dataforgood.fr/

Data.gouv.fr - FR open data potal, https://www.data.gouv.fr/

Data.gov.uk - UK open data potal, https://data.gov.uk/

Dataportalen - beta, https://dataut.vegvesen.no/

Data-story #1 Handisco (Nancy), https://www.data.gouv.fr/fr/posts/data-story-1-handisco-nancy/

Directorate for Administration and ICT (Difi), https://www.difi.no/

ENTUR - Entur-appen, https://www.entur.org/entur-appen/

Entur – Journey Planner, https://play.google.com/store/apps/details?id=no.entur&hl=en

Entur AS app info. of website, https://www.entur.org/

EU Open Data Portal - Access to European Union open data, https://data.europa.eu/euodp/en/home

European Commission (2014), guidelines on recommended standard licences datasets and charging for the reuse of documents. https://ec.europa.eu/digital-single-market/en/news/commission-notice-guidelines-recommended-standard-licences-datasets-and-charging-re-use

European Data Portal (2015), Creating Value through Open Data, Study on the Impact of Re-use of Public Data Resources, https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

European Data Portal (2017), Re-Using Open Data, A Study on Companies Transforming Open Data into Economic and Societal Value, January 2017. https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf

Gender Equality, http://eufunding.justice.ie/en/eufunding/pages/genderequality

Geofence as zero emission zone - Statens Vegvesen, https://www.vegvesen.no/fag/trafikk/its/geofence-til-nullutslippssone-demo

Global Report 4th Ed., https://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf

Immonen, A., Palviainen, M., & Ovaska, E. (2014). Towards open data based business: Survey on usage of open data in digital services. International Journal of Research in Business and Technology, 4(1), 286–295

LeMO - Report on Legal Issues, https://lemo-h2020.eu/newsroom/2018/8/31/deliverable-21-report-on-economic-and-political-issues

LeMO: Report on Legal Issues, https://lemo-h2020.eu/newsroom/2018/11/1/deliverable-d22-report-on-legal-issues

Ministry of Climate and Environment, http://luftkvalitet.info/home.aspx

Network Timetable Exchange - Standards Context: Transport Standard Context, http://netex-cen.eu/?page_id=58

Norwegian Institute for Air Research (NILU), https://www.nilu.no/

Norwegian Licence for Open Government Data (NLOD) 2.0, https://data.norge.no/nlod/en/2.0

Open Data Barometer, http://devodb.wpengine.com/

Open Data Institute (2016a). What makes data open? Retrieved from: http://theodi.org/guides/what-opendata

Open Data Maturity in Europe - Report 2018 by EUROPEAN Data Portal, https://www.europeandataportal.eu/en/news/open-data-maturity-europe-2018

Open Data Maturity in Europe, 2018, https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n4_2018.pdf

Open public data in Norway - Difi, https://data.norge.no/

ORR office of rail and road, http://dataportal.orr.gov.uk/

Personal data in transport: exploring a framework for the future - Open Data Institute (ODI), https://theodi.org/article/personal-data-in-transport-exploring-a-framework-for-the-future-report/

Re-Using Open Data, A Study on Companies Transforming Open Data into Economic and Societal Value, 2017, https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf

Social Inclusion and Community Activation Programme (SICAP) 2018 – 2022, https://www.pobal.ie/programmes/social-inclusion-and-community-activation-programme-sicap-2018-2022/

Statens vegvesen – open data APIs, https://www.vegvesen.no/om+statens+vegvesen/om+organisasjonen/apne-data

Tammisto, Y., & Lindman, J. (2012). Definition of open data services in software business. In Software Business (pp. 297–303). Springer.

The National Road Database - Statens Vegvesen, https://www.vegvesen.no/en/professional/roads/national-road-database

Transport data in the UK and France - Open Data Institute (ODI), https://theodi.org/article/transport-data-in-the-uk-and-france-report/

Trusted Data Services for Global Science , http://www.icsu-wds.org/organisation

TURBLOG. (2011). FP7 TURBLOG Deliverable 2 Business Concepts and Models for urban logistics.

UK Digital Strategy, 2017, https://www.gov.uk/government/publications/uk-digital-strategy

UK Government Transformation Strategy 2017 to 2020, https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020

UK Open Government Licence for public sector information, https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

Valencia - La NUEVA APP,https://www.valencia.es/ayuntamiento/atencion_ciudadano.nsf/vDocumentosTituloAux/Aplicaciones%20m%C3%B3viles?opendocument&lang=1

Verhulst, S. & Caplan, R. (2015). Open Data. A twenty-first-century asset for small and medium sized enterprises. GovLab, April 2015.

Zuiderwijk, A., Janssen, M., Poulis, K., & van de Kaa, G. (2015). Open data for competitive advantage: insights from open data use by companies (pp. 79–88). ACM Press. doi: 10.1145/2757401.2757411.

5 Case study 3: "Real-time traffic management"

The case study on "Real-time traffic management" focused on the use of big data in the management of road traffic management systems and strategies. Interviews were carried out with the Tallinn city transport department, the Research Centre of Logistics and Transportation of the Tallinn University of Technology, and Entur A/S, the company owned by The Norwegian Ministry of Transport and Communications in Norway.

5.1 CASE STUDY OVERVIEW

5.1.1 Organisation/Stakeholders

The relevant stakeholders of the traffic management business consist of a series of actors playing various roles, for instance:

- **Government body/policy marker & regulation:** has a role for defining the policies and monitoring the compliance with regulations and legislations related to the services.
    - *Country/Ministry/Province/Municipality/Sub-city*: the hierarchy of public administration for defining policies, providing grants and funding Intelligent Transport Systems (ITSs) services.
    - *Traffic Manager/Road Operator*: supervises the traffic management of an area (e.g., a city or a highway) and is responsible for its optimization.
    - *Police/Enforcement*: monitors violations of the "Code of the Road" and related law. It includes public safety answering point, which is the collection centre for emergency calls and rescue.
    - *Certification body*: certificates the adherence and compliance of products and services with standards and technical guidelines.
    - *Public Transport*: plays a role of train/metro/bus/tram Operators.
    - *Auxiliary body*: includes ambulances or fire brands that in some cases require priority on a road.
- **Traffic service provider:** is contractually providing services directly or indirectly from the producers to consumers. Usually a software/technological company or a navigation provider acts as a service provider.
- **User/consumer:** indicates the stakeholders who are perceived as users of the service (public, commercial or private) and who are willing to pay the provider for services.
    - *Individual road user*: means a driver or a vulnerable road user (e.g., a pedestrian or a cyclist) as end users on roads.
    - *Transport Company*: are fleet managers, actors who manage a number of vehicles, such as busses, emergency vehicles, trucks or taxi and truck drivers.
- **Technology supplier**: is supporting the producers of the functionality of the service(s) or the service provider with the necessary technology and devices. It can be any actor providing devices, hardware platforms, software applications, consulting services to all the other actors involved in the services like:
    - *Road Side Unit (RSU) provider*: provides complete RSUs and in some cases performs the task to install and maintain the RSUs in the road infrastructure.

 - *Road Sensor provider*: produces any type of sensor (e.g., camera, speed sensor, location module and actuators) to be connected or integrated in a RSU in order to capture real data and information.
 - *On-Board Unit (OBU) provider*: provides the OBUs to the car/truck maker or retrofit installer in aftermarket scenarios.
 - *Vehicle maker*: plays the role of a maker of every kind of vehicle (cars, trucks, buses, ambulances, fire-fighters vehicles, etc.).
 - *IT provider*: supports IT hardware and software for Back-Office operations.
- **Service enabler**: is supporting service providers with necessary services and contents.
 - *Content provider*: finds and creates content (traffic data, information or basic services) to build useful services (e.g., POI on maps) for end users.
 - *Connectivity provider*: provides the SIM card/module to be inserted into the OBU and RSU, connectivity services to users and other actors, other value-added services like location or identity management.
 - *Broker*: is a facilitator between service sellers and consumers by providing an integrated platform for mobility services.
- **Event location provider**: owns or hires facilities for event hosting.
- **Event organizer**: organizes events and eventually gathers people in a specific location
- **Parking operator:** is a stakeholder that owns, manages and provides parking facilities and services. It can be either public actors, private business actors or both.

The ecosystem of traffic management market involves various actors, from providers to consumers, classified as shown in the figure below.

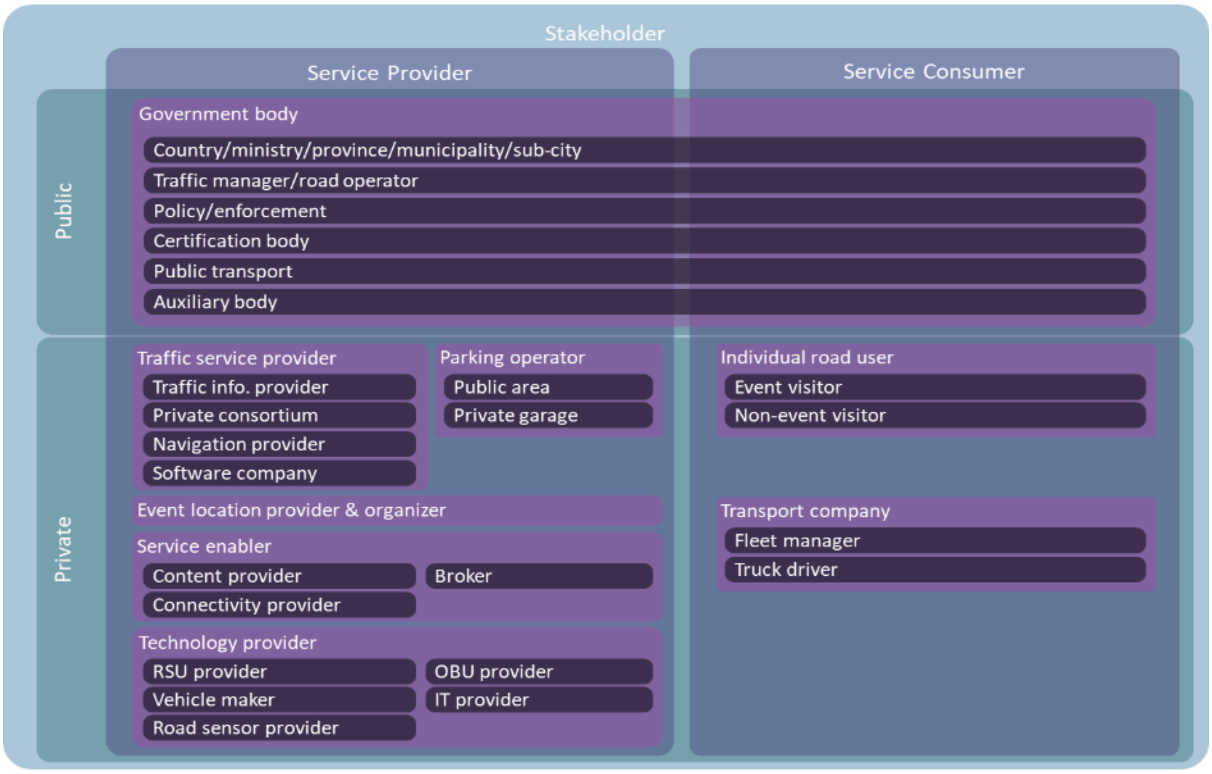

Figure 12 Taxonomy of stakeholders related to traffic management in transport sector

Table 11 Interviewees of the case study

Organisation/Type	Designation	Knowledge	Position/Role	Interest
Tallinn transport department (Government institution) / National body (Estonia)	Head Chief specialist Development specialist	Multi-modal transport	Acquisition Storage Usage	Road maintenance Traffic management
Entur AS (Government institution) / National body (Norway)	CTO Data analyst	Multi-modal transport	Analysis Curation Storage Usage	Improve multi-modal mobility service Increase access to public transport data
Tallinn University of Technology, research Center of Logistics and Transportation (Research institution) / Province (Estonia)	Professor (project manager)	Multi-modal transport	Analysis Usage	Real-time traffic management

5.1.2 Sector/Mode

This case study deals with **real-time traffic management** of road transport using big data. As such, all road users from cyclists to truck drivers are affected by its developments.

5.1.3 Case study motto

This case study investigates the use of big data in the context of increasingly sophisticated intelligent transport systems.

Big data allows warnings ahead of traffic jams, lane handling, traffic flow predictions, etc. Going from Big Data to active traffic management requires merging Big Data with data from fixed sources. The use of archived data allows improving individual route planning, to measure bottlenecks and delays, to measure system reliability, to determine priorities for infrastructure improvement, and to analyse the impact of the investments made. Moreover, without accurate demand estimation, it is difficult for transport operators to provide their services and make other important traffic management decisions. These benefits from Big Data have been further emphasised with the advance of (cooperative) intelligent transport systems, which is expected as a key enabler for future traffic management systems. In this case study, we will investigate and discuss challenges and opportunities of real-time traffic management that is an important part of currently desired intelligent transport system in transport sector with respect to various perspectives such as technological, economic, legal, policy, societal, ethical and political aspects.

5.1.4 Executive summary

Traffic congestion has become increasingly serious, and traffic accidents have occurred frequently in many cities due to urbanization and increasing vehicle demand. These have become traffic management problems that need to be solved. In response to the increasing traffic demand and the pressure of transportation resources, the traditional way of traffic management is showing itself inadequate[47], such that many countries have tried to introduce Intelligent Transport Systems (ITS) to manage the most important sections of traffic. By use of modern information technology, based on Big Data, the traditional road traffic management methods are deeply reformed to improve the efficiency of urban traffic network, ease urban traffic problems, reduce unnecessary losses and improve the efficiency of public transportation[48]. At present, most cities in developed countries have completed or are building complete three-dimensional traffic monitoring platforms. Vehicle locations, pedestrian locations, camera and other monitoring tools are used to comprehensively monitor vehicle travel speed, road cross-section flow, and intersection shunt. By use of these data in timely, real-time evaluation of the traffic status of roads might be possible. Huge and real-time monitoring data forms massive traffic information, which provides effective data support for traffic congestion prediction, and also requires traffic congestion prediction to efficiently and comprehensively cover the entire urban transportation network[49].

Big data allows warnings ahead of traffic jams, lane handling, traffic flow predictions, etc. Going from Big Data to active traffic management requires merging Big Data with data from fixed sources. The use of archived data allows improving individual route planning, to measure bottlenecks and delays, to measure system reliability, to determine priorities for infrastructure improvement, and to analyse the impact of the investments made. Moreover, without accurate demand estimation, it is difficult for transport operators to provide their services and make other important traffic management decisions. These benefits from Big Data have been further emphasised with the advance of (cooperative) intelligent transport system, which is expected to be a key enabler for future traffic management systems.

In this case study, we will investigate and discuss challenges and opportunities of real-time traffic management emphasizing the use of big data in ITS with respect to technological, economic, legal, policy, societal, ethical and political perspectives.

---

[47] Wences, Pedro, et al. "Decision-Making Intelligent System for Passenger of Urban Transports." International Conference on Ubiquitous Computing and Ambient Intelligence. Springer, Cham, 2017.

[48] Guan, Shuqi, et al. "Intelligent transportation system contributions to the operating efficiency of Urban traffic." Journal of Intelligent & Fuzzy Systems 31.4 (2016): 2213-2220.

[49] Wang, Zhenhua, Yangsen Yu, and Dangchen Ju. "Analysis and Prediction of Urban Traffic Congestion Based on Big Data." International Journal on Data Science and Technology 4.3 (2018): 100.

5.2 TECHNICAL PERSPECTIVE

5.2.1 Data sources/Uses

General traffic information means information derived from any road and traffic data, or their combination thereof, provided by any road authorities, road operators or service providers to road users through usual communication channels. Real time traffic information relates to current traffic conditions on the road network. Such information includes for instance accident locations, incident warnings (incl. safety related events / conditions), road works, congestion hotspots, travel times / delays. Such services fall within the scope of delegated Regulations (EU) 2013/886[50] & 2015/962[51].

Transport traffic related datasets can be categorised into the following groups: real time road weather data, travel times, web camera images, incidents, road works, driving conditions and road closures & diversions. More data detail groups with its usages and dimensions of Big Data are listed into *Table 12*.

According to EU EIP SA46 Annual NAP report - 2016[52], in total 12 member States have responded to the survey about the current status of implementation of the National Access Point (NAP[53]) for the provision of EU-wide real-time traffic information services, in short 'NAP for real-time traffic information'. The Commission Delegated Regulation on this topic ((EU) 2015/962) adopted in 2015 applies as from 13 July 2017, so at the time of the survey there was no obligation yet to have a NAP on Real-Time Traffic Information (RTTI) up and running. Only three countries (i.e., Cyprus, Finland and Germany) have a (partly) operational NAP for RTTI. Seven other countries (Austria, Denmark, The Netherlands, Norway, Portugal, Sweden and the UK) have concrete plans to implement an NAP. In Belgium and Poland the NAP for RTTI is not yet operational or planned. These RTTI can be grouped as presented in *Table 13*.

Table 12 Main data sources (traffic information) of the case study

Data source	Used in	Big Data dimensions	Other remarks
Events and traffic incidents	Dynamic speed management Prohibition of truck take-over Implementation of reversible lanes	Volume, Velocity, Veracity, Variety	From Variable Message Signs (VMS), web servers, RDS-TMC[54], information telephones, SMS, call, Teletext and interactive digital televisions or smartphone Apps.

[50] Commission Delegated Regulation (EU) No 886/2013, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex% 3A32013R0886

[51] Commission Delegated Regulation (EU) No 962/2015, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex% 3A32015R0962

[52] EU EIP SA46 Annual NAP report – 2016, https://www.its-platform.eu/filedepot_download/1971/6019

[53] Intelligent transport systems – National Access Point, https://ec.europa.eu/transport/themes/its/road/action_plan /nap_en

[54] RDS-TMC: Radio Data System – Traffic Message Channel

<table>
<tr><td>Traffic flow (Levels of service)</td><td rowspan="8">Management of high-occupancy lanes
Ramp metering
Dynamic management of driving restrictions in mass movements and adverse weather conditions
Tunnel management
Traffic management plans
Dynamic management of urban traffic plans
Traffic lights priority systems for the public transport
On request public transport
Public bicycles services management
Car-pooling and car-sharing</td><td>Volume, Velocity, Veracity</td><td>From private operators (Google, INRIX, etc.) and toll motorway operators.</td></tr>
<tr><td>Travel times</td><td>Volume, Value</td><td>From VMS, web servers, smartphone Apps by administration, municipalities or service providers on road networks.</td></tr>
<tr><td>Information of speed limits</td><td>Value, not dynamic</td><td>From VMS, web servers or smartphone Apps by relevant administrations or toll operators on all road networks.</td></tr>
<tr><td>Driving restrictions</td><td>Value, not dynamic</td><td>From VMS, web servers, channels of news, smartphone Apps by relevant administrations.</td></tr>
<tr><td>Images or video distribution</td><td>Volume, Veracity, Variety</td><td>From web servers or smartphone Apps of service providers on big cities and interurban.</td></tr>
<tr><td>Weather-related information</td><td>Volume, Velocity, Veracity</td><td>From web servers or smartphone Apps of service providers on interurban network.</td></tr>
<tr><td>Itinerary planning</td><td>Volume, Velocity, Value</td><td>From web servers or smartphone Apps of service providers on road networks.</td></tr>
<tr><td>Information exchange</td><td>Volume</td><td>In formats of DATEX, XML, text.</td></tr>
</table>

Table 13 Real-time traffic information by extend of dynamics

Periodical processing	(Near or) Real-time processing	
Static road data	Dynamic road status data	Traffic data
Road network links and physical attributes (e.g. geometry, road width, number of lanes, gradients, junctions)	Road / lane / bridge closures, Accidents, Incidents	Traffic volume
Road classification	Overtaking bans on HGV	Speed
Traffic signs on traffic regulations and dangers(e.g. access conditions for tunnels / bridges, permanent access restrictions, other)	Road works, Poor pavement conditions	Location and length of queues, Travel times
Speed limits	Dynamic speed limits	Waiting time at border crossings to non-EU countries
Freight delivery regulations, Traffic circulation plans	Temporary traffic management measures	
Location of tolling stations	Direction of travel on reversible lanes	

Tolled roads, fixed RUC, payment methods	Variable RUC, payment methods	
Location of parking places / service areas	Availability of parking places, cost of parking	
Location of charging points for EV and conditions of use	Availability of charging points for EV	
Location of CNG / LNG / LPG stations	Availability of delivery areas	
Location of public transport stops and interchange points	Weather conditions affecting road surface and visibility	
Location of delivery areas		

In Norway, existing real time information services, both static and dynamic, include the entire national road network. Trans-European Transport Network (TEN-T[55]) roads and motorways have the highest priority for instrumentation. This coverage also applies for the implementation of Delegated Act (EU) 2015/962[56]. The Norwegian Public Roads Administration (NPRA) has had a national traffic information service in operation since the early 1990's. The service is based on five regional Traffic Management Centres (TMCs). Dynamic data on driving conditions, traffic, travel times, road works and incidents are either automatically collected from roadside units (traffic and road weather stations, tolling roadside units, cameras etc.) are reported to the TMCs by the police, contractors, media or road users. Information regarding incidents, road works, restrictions and closures are registered in the national Traffic Information Centre (TIC) system.

In addition to these, NPRA is also responsible for the National Road Data Base (NRDB). This database includes the Norwegian road network such as national, county and municipal roads as well as private roads and stores both basic data and statistical data like traffic accidents and average annual daily traffic. In addition to these data, NRDB contains basic data on airports and all stop points and terminals for bus, subway, railway and ferries. Other static data are road geometry, driving restrictions, traffic plans, permitted axle load, surface material, road curvature, speed limits, road width, tunnels and bridges, road furniture, rails, traffic signs, manholes, ditches, brick walls etc.

Environmental data are also registered into the database. The database offers a standardised web application based on the standard REST-API, for retrieving road and road traffic data. Data types will be road network geometry, road characteristics, traffic regulations, speed limits and many other types relevant for digital maps and information systems such as travel planners and navigations services. The data are able to be used by service providers and digital map

[55] Trans-European Transport Network (TEN-T), A European Commission policy directed towards the implementation and development of a Europe-wide network of roads, railway lines, inland waterways, maritime shipping routes, ports, airports and rail-road terminals. It consists of two planning layers: comprehensive and core networks, https://ec.europa.eu/transport/themes/infrastructure/about-ten-t_en

[56] Commission Delegated Regulation (EU) 2015/962, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32 015R0962

producers. This database offers an API for retrieving traffic data statistics free of charge. The real time traffic data for volume, speed and vehicle categories from the registration points will shortly be made available in the NPRA Datex-node and therefore also the National Access Point (NAP[57]).

As a national access point for travel and public transport information in Norway, Entur collects all schedule data and real-time traffic data for public transport and stops, and shares these data to Google, Apple, NSB, Kringom and etc. through developing APIs and standardizing (NeTEx: European standards[58]). NeTEx is the CEN (European Committee for Standardization[59]) technical standard for exchanging public transport information as XML documents. It provides a W3C XML schema based on the Transmodel[60] abstract model of common public transport concepts and data structures. This schema can be used to exchange many different kinds of data between passenger information systems, including data describing for stops, facilities, timetabling and fares. Such data are used by both operational management systems and customer facing systems for journey planning etc. Also, Entur gathers traffic data by considering Action A of NAP having quite detailed regulation of what kind of data have to be available, when, and what format, and so on from the directive. Therefore, most of public transport is covered by this directive.

Tallinn is the capital of Estonia and with 400,000 inhabitants and an area of 159.2 km² the country's largest city. Since Estonia gained independence in 1991, Tallinn has undergone drastic change, which the public transport system could not keep up with. The city is struggling with more and more traffic from private cars. Tallinn has developed a traffic development plan for 2005 - 2014 to renew and extend the public transport network including buses, trolleys, tramways and suburban trains. In this context, Transport Department of Tallinn (TDT) has many tasks related real-time traffic management such as guaranteeing traffic and coordinating the planning, establishment, maintenance and use of objects of public transport infrastructure and taking measures for creating favourable traffic conditions for means of public transport. They are managing traffic based on various types of information (e.g., timetables, public transport map, tickets, parking, traffic cameras, bicycle, taxi, airport, port, heavy transport)[61].

5.2.2 Data flows

There are five key mechanisms for data creation (**manual collection, overt crowd-sourcing, covert crowd-sourcing, sensor-derived, service provider generated**). Web-connected fixed and mobile sensors, plus crowd-sourcing are the emerging transport data collection mechanisms. In particular, personal/vehicle location data are important for intelligent mobility.

In general, as shown in the figure below, traffic management systems, which consist of five layers, interact with various actors as external entities. The main categories are:

---

[57] National Access Point (NAP), https://ec.europa.eu/transport/themes/its/road/action_plan/nap_en
[58] Wikipedia – NeTEx, https://en.wikipedia.org/wiki/NeTEx
[59] Wikipedia – CEN, https://en.wikipedia.org/wiki/European_Committee_for_Standardization
[60] EU regulatory and use of Transmodel, http://www.transmodel-cen.eu/overview/use-of-the-transmodel/
[61] Transport information of Tallinn, https://www.tallinn.ee/eng/otsing?sona=11964

- **Vehicle Driver**: An actor driving in a vehicle. The vehicle is a motorized vehicle (car, bus, truck) and not a vehicle of a vulnerable road user (bike, moped, motor).
- **Vulnerable Road User (VRU)**: A VRU is a human actor like a pedestrian, cyclist or powered two-wheel driver (PTW); a motorcyclist is also an example of a PTW and is treated as a vulnerable road user in specific road hazard situations with other cars.
- **End User**: A human actor who uses a product or service as an individual, i.e. not on behalf of an organisation.
- **Road Operator**: An actor responsible for the traffic management of a road network.
- **Service Provider**: An actor (organisation) supplying services to one or more customers. Customers are either other organisations, including government (B2B / B2G / G2B / G2G) or end users (B2C / G2C). Typical examples of service providers are a Navigation Provider or a Traffic Information Provider.
- **Other**: Any other actors (e.g., event organizers) that have interaction with the system or is interested in the deployed applications.

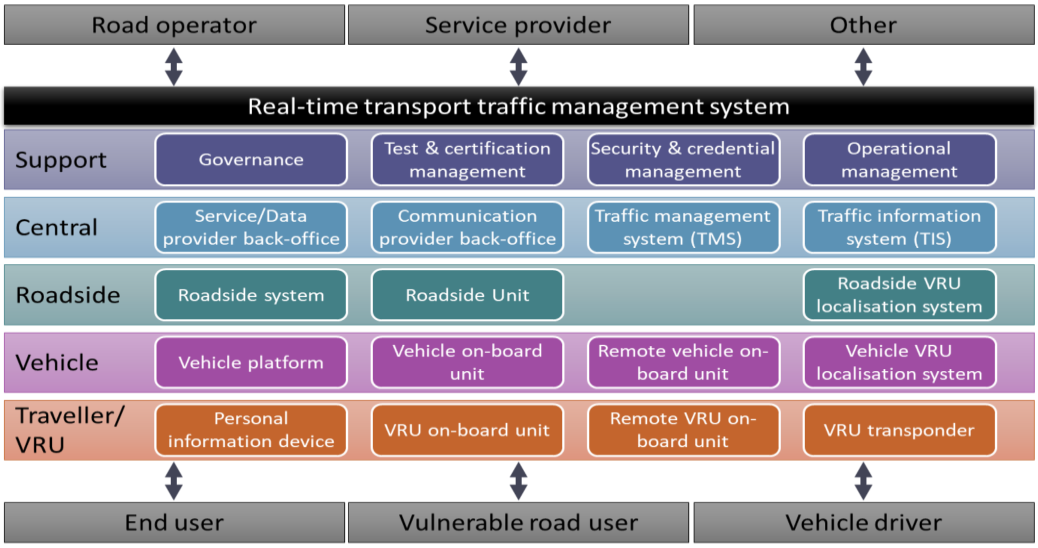

Figure 13 Data flow and components of real-time traffic management

As mentioned in the section on Case study 2: "Open data and the transport sector", the role of private sectors, which play a growing role in collecting and aggregating transport-related data, drives the availability of personal/vehicle location datasets. However, the increase of partial datasets created by the private sectors causes problems with the interoperability of data and the controlling of data, since we know less about what data being gathered represent and where data are going to. Test, certification and security management become very important, but few public sectors have resources to do this. For example, Norwegian public Roads Administration (NRPA), which is operating five regional traffic management centres, is leading the way as a key player, while many local authorities are still developing integrated digital strategies. In addition, datasets collected in real-time are mainly processed by private sectors (Google, INRIX and so on) and are not often shared with the public sector in order to develop public services to the benefit of society under governance management. This still causes difficulties when integrating Traffic Management Systems (TMS) into a Single Market and finding proper datasets for the provision of real-time management, although EU action plans such as national access point are trying to consolidate various TMS industries into the Single market.

5.2.3 Related big data policies to traffic management

The 2017 ITS national reports demonstrates the strong and constant involvement of most EU countries in intelligent traffic management and information systems[62]. This allows for a better use of the infrastructure especially via better use of road, traffic and travel data and the development of new intelligent transport services for traffic management. In addition to these, open data strategies for transport (e.g. in the UK) or the use of crowd-sourcing (e.g. travel-time information in Finland) have led to significant changes and new services. In the context that most of traffic management tasks have been included into ITS nowadays, we look at the below figure which shows ITS action plans published by European Commission. The plans related to traffic management are mainly located to the Area 2 of the action plans.

According to the report named "updating the Working Programme in relation to the actions under Article 6(3) of Directive 2010/40/EU[63]", action plan for EU-wide real-time traffic information services included into priority area I of the ITS Directive has been revised for specific data types, the possible geographical extension of its current specifications and possible additional data types, especially relevant data types at urban level. It aims to extend the geographical scope so as possibly to cover the whole road transport network, at least for some data types to be determined, and to look at possible new data types such as urban vehicle access restrictions by 2019.

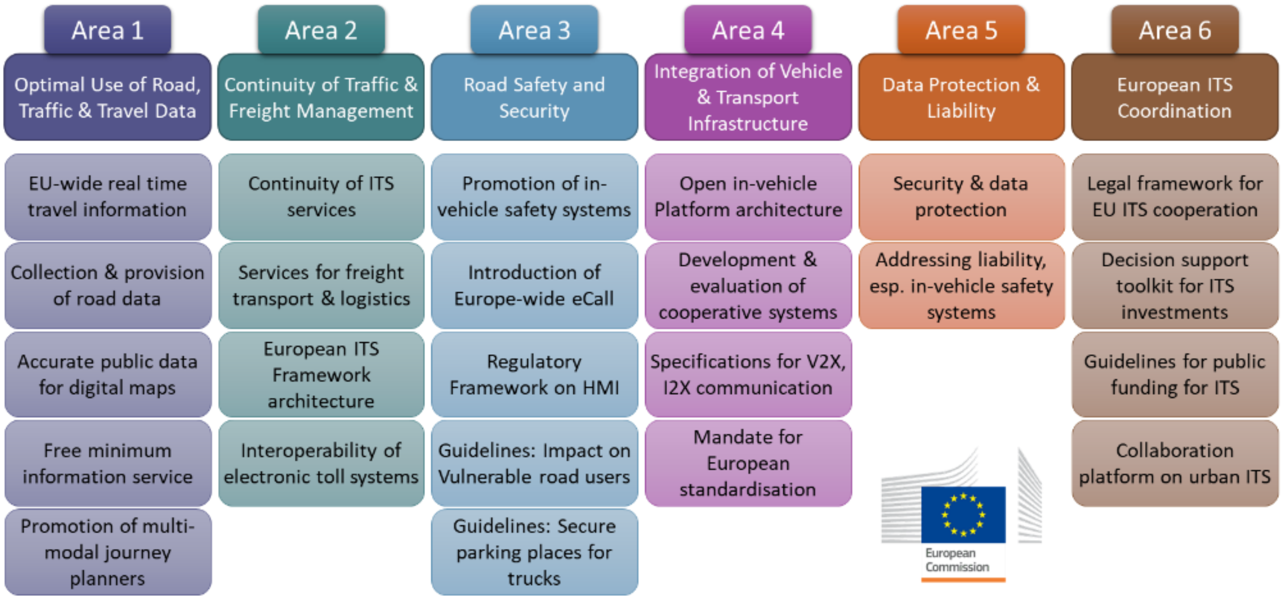

Figure 14 ITS action plan - European Commission

Access to vehicle data for road operation purposes in action plan of priority area I of the ITS Directive is considering on access to vehicle data for the needs of public authorities, road operators and any other parties in charge of road operations, especially for traffic management

[62] ITS national reports, https://ec.europa.eu/transport/themes/its/road/action_plan/its_national_reports_en

[63] updating the Working Programme in relation to the actions under Article 6(3) of Directive 2010/40/EU, 2018, https://ec.europa.eu/transport/sites/transport/files/legislation/c20188264_en.pdf

purposes. Its objective is to work on the data needs and the roles of parties in the business-to-government context, taking into account current commercial activities by 2019.

For continuity of traffic and freight management services, action plan in priority area II of the ITS Directive has being additionally developed. Several initiatives already address actions in this area. Its purpose is to assess the need for complementary actions such as standardised information flows or traffic interfaces between traffic information/control centres and various stakeholders. Appropriate interfaces and interactions between the traffic management and information systems from the various stakeholders (e.g., road operators, vehicle manufacturers and service providers) should be ensured in order for them all to contribute to safer and more efficient traffic flow. The mapping exercise with Member States experts will take into account the recommendations of the C-ITS platform, in particular those relating to enhanced traffic management in 2019 to 2022.

The state-of-the-art in Network and Traffic Management is facing a deep transformation, in order to tackle the upcoming challenging objectives and to deal with the system constraints. This transformation is made possible by three major evolving factors (TEN-T and CEF regulations[64]): improvement of the existing infrastructure and facilities, gradual introduction of new generation vehicles, and the development of new technological systems, governance and procedures to better manage traffic operations and to offer new types of passenger and freight services. In accordance with the TEN-T and CEF regulations, the current infrastructure investments are concentrated in construction, upgrade, and modernisation of the infrastructure, in order to enable improved interoperability and enhanced efficiency. This is made possible by improving cross-border sections, removing the existing bottlenecks and bridging missing links. For example, the removal of level crossings is a key undergoing action to alleviate traffic on both road and rail lines and to reduce the risk of accidents. In road traffic, efforts are currently dedicated to improve the capacity of the busiest arteries and of the highly dense urban environments.

5.2.4 Main technological challenges and issues

The transport sector, especially traffic and infrastructure management, is facing a deep transformation related to the challenges and to the introduction of new services for all types of users, with the aim of maximising the benefits and minimising the costs.

After data collection and analysis, the knowledge acquired from it needs to be represented in a correct **way to represent real traffic condition**, otherwise it may occur false positives or wrong information. Regarding this, one of key challenges is how to converge many different types of information into a single traffic condition representation. It means which information has more or less importance to the traffic and how each one will impact traffic. Providing such representation is still a big issue. The interview with Entur AS has emphasized that data representation is also important in real industry to move from simple traffic management to intelligent transport systems.

[64] CEF transport projects by country can be find by https://ec.europa.eu/inea/connecting-europe-facility/cef-transport/projects-by-country

In addition, many traffic management systems have been proposed to support such representation in order to **identify traffic hazards**. However, many of them are inefficient or they cannot identify such hazards as soon as it occurs. Thereby, it uses predefined intervals to try to identify these hazards, since exploiting all relevant traffic data is complex or impossible. However, which interval is the best to try this identification? Since with too small interval, the traffic management system may not receive enough information to identify the hazards. Otherwise, a slow performance of the traffic management system may occur via a large processing of data within a great interval.

Guiding and computing alternative routes to avoid traffic hazards are the better options to improve the overall traffic efficiency. However, the main challenge is how to do the process in an acceptable time to not to introduce an undesired overhead. Although relying on central entities (centralized approach) to compute and suggest alternative routes to all vehicles is more efficient due to its better management and scenario overview, it depends on the number of vehicles to be re-routed and the complexity of the algorithm used in the alternative route computation. If with the huge number of vehicles and complex algorithm, it may introduce high overhead, degrading its performance. In this regard, one solution is to enable each vehicle to compute its own alternative route. However, to enable every vehicle to compute an efficient route based on a full scenario overview about the traffic condition could cause the overloading of network. Another concern is how to compute an efficient alternative route without incurring in traffic congestion in other areas in the nearby future, providing a better traffic balance and management. In this way, to have a good alternative route guidance, a trade-off between efficiency and complexity is essential.

5.2.5 Big data dimensions and assessment

In terms of variety of Big Data, despite the traffic management systems enabling the data integration with different sources to improve its overall performance, **heterogeneous data integration** is still a big issue. Since we have lots of different systems and sources generating a huge amount of various data, main challenge is how to do the integration, without unified **standardization**. Furthermore, as emerging technologies such as IoT will provide **data exchange and communication** to a plethora of everyday life devices, it is important to use these devices to turn the data collection paradigm into a new one. However, with this integration, many other challenges will arise including **tracking and managing** data from the high number of devices that will be involved in such integration.

Regarding **Big Data management issues**, traffic management systems are responsible for handling a huge amount of data (volume), and many sources may report its data asynchronously in different time levels (different velocity). Furthermore, data correlation is another challenge due to non-integration among different systems and sources, in which the same source may provide data in different systems. In other words, as different systems are independent, the data accounting can incur in false positives (veracity). However, the major challenge is how to exploit these Big Data issues in a vehicular environment, once the current models and algorithms used in Big Data are physically and logically decentralized, but virtually centralized.

The next frontier for innovation and productivity of traffic management is the capability to perform **real-time data processing and analytic**, with respect to volume and velocity of Big

Data. Most of existing ITS 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). This leads to a complex paradox, since for most transportation and mobility problems a reduced response time is critical to ensure that information and decision making are tightly coupled in time and thus, useful in practice. In addition, real-time processing is deemed crucial to provide new forms of on-demand mobility for people and things, requiring real-time data sharing and analysis, calling for legislative/public attitude shifts to implement intelligent traffic management systems using various data. Summarising, real-time service automation and optimisation is a key enabler for intelligent traffic management systems in the future.

5.3 BUSINESS MODEL

The figure below presents a business model of traffic managers and operators as authorities of government for traffic management in the form of business model canvas[65].

Key partners	Key activities	Value proposition	Customer relationships	Customer segments
Who are the firm's partners? Public sectors at national, local, regional levels. Private sectors from parking operators, event providers, service enablers, transport companies, traffic service providers and technology providers.	**What are the key activities, such that the firm operates successfully?** Data collection, validation, aggregation, analysis and management of traffic at region and national levels	**What value does the firm deliver to the customer?** Data services and products based on traffic data by collaborating with partners	**What type of relationship is established between the firm and the customer?** Collaboration by relevant directives, policies, initiatives, contracts and standards	**For whom is the firm creating value?** Various targets such as most of collaborators for end users (i.e., citizens)
	Key resources	**Internalization of costs and benefits**	**Channels**	
	What does the firm need to create value? Proper techniques for big traffic data analysis and decision making, policy and legal supporting for adopting these techniques	**How are costs and benefits from firm's business activities internalized?** Increasing the number of jobs and human resources (citizens) for advanced traffic management by using budget from government and municipalities. Generating new goods and services of collaborators using traffic data and information for efficient public services and relevant market growth.	**How does the firm reach the customer?** Various communication channels (email, SMS, SNS, call and mobile apps)	
Cost structure			**Revenue streams**	
What are the costs associated to the business model? It depends on budget capability of each municipality and authorities.			**For what value are the customers paying and how are they charged?** Feedback collected by collaborators from end users to improve the performance of traffic management and to reduce traffic congestion.	

Figure 15 Business model for traffic managers and operators

The business could collaborate with various actors from public and private sectors, and at national, local and regional levels, as mentioned in Section 1 and 5. Their key activities vary such as data collection, validation, aggregation and analysis related to traffic management, according to data resources relevant, techniques needed, policy and regulation supporting to adopt proper technologies. Some of actors have both roles as customer and provider, in other words, customers are able to be the relevant companies, service providers and citizen. Therefore, other companies might be customers and collaborators each other based on relevant directive, policy, initiative, contract, and license or data ownership. To gather feedback from end users (citizen), the companies or service producers connected with the users by their products and services receive the opinions via email, SMS, SNS, call and mobile apps, and deliver that to upper providers such as traffic managers and operators, public transport and so on. Of course, these

[65] TURBLOG. (2011). FP7 TURBLOG Deliverable 2 Business Concepts and Models for urban logistics.

providers also have the communication functions to get the information of real situations (e.g., accident) on roads. The feedback contributes to creating revenues and increasing the number of jobs involved in producing the service or product and save costs, related to road traffic. These benefits lead to increase utilisation of their goods and services using traffic data, enhance customers' loyalty and grow relevant market.

As a concrete example, business model of Entur and Tallinn Transport Department in accordance with traffic management is described as follows:

- *Value proposition*. These two government institutions collect relevant traffic data to maintain roads; manage traffic in near- or real-time; improve multi-modal mobility services.
- *Customer segment*. Range of target customers is very broad from relevant transport providers, but main customers are people using ticketing services for public transports and road users driving their own vehicles.
- *Customer relationships*. As non-profit authorities, both of them do not have specific competitors in regional, national and global levels. Instead, there are collaborations with other relevant institutions such as transport companies, traffic service providers, individual road users, parking operators and event providers & organisers to manage effectively and efficiently traffic. For example, TDT cooperates with Ridango (technology provider), Disainiosakond (service provider), Port of Tallinn (other authority) and so on.
- *Channel.* Various methods such as email, SMS, call, chatting and apps to communicate with their customers are used. Most of feedback from the customers is related to traffic, since it directly impact to them. For example, TDT receives 500,000 feedbacks about traffic signal light per year through Tallinn city webpage[66], however stakeholders in the department expect that the number of feedback will be decreased according to leveraging real-time traffic management system in future. In addition, parking related feedback is increasing nowadays, since parking systems can be easily connected by mobile device of users.
- *Key activities.* Key activities to make the value proposition are gathering traffic data and data associated with transport; analysing the data; leveraging the data to improve traffic management. For example, TDT is immediately providing traffic information (e.g., closed-off roads and streets, traffic rearrangement and detour routes)and video from CCTV on roads in (near) real-time, and applications for installing means for traffic information are gathered by them to maintain road and manage traffic congestion.
- *Key resources.* A lot of human resources are required for maintaining road infrastructure and managing traffic. In particular even real-time traffic management in which most of their functions operate automatically requires the resources to maintain sensors and update road status by report from people. It could be complementary by collaborations with third parties dealing with these tasks, e.g. Entur has their own customer call centres and collaborates with other transport operators and service providers.
- *Key partners.* Key partners are from transport providers to event planners and operators. In the case of TDT, institutions which are dealing with various transport

[66] Transport Department of Tallinn – Reception of complaints related to traffic control devices, https://www.tallinn.ee/Teenus-Reception-of-complaints-releated-to-traffic-control-devices

modes such as bicycle, taxi, airport, pot, freight and parking are main partners to manage traffic.

- *Cost structure.* All costs are funded by government. However, Entur earns additional profits by providing premium services to cover the cost of human resources needed to interact with end-users.
- *Revenue streams.* Even though no direct compensation is paid for traffic management, managing traffic has indirect benefits such as improving traffic management, reducing air pollution and saving energy.
- *Internalisation of costs and benefits.* As NAPs in Norway and Estonia, all external costs are funded by government, therefore, the benefits of traffic management is returned for society.

5.3.1 Business processes

Traffic observance is a rising trend and is growing with the massive demand on the traffic sensors market. Traffic observance uses traffic information sensors and management systems for vehicle classification, shaft investigating, over height, and 3D vehicle identification. Traffic management systems have become the most popular segment in the market and are used for managing traffic in cities and minimizing traffic jams on roads. Recently, prominent traffic management solutions include various systems related to ITS such as smart signalling systems, traditional signalling and video surveillance systems, route guidance systems, parking systems and intelligent video management systems. The demand for traffic management market is driven by factors, such as high demographic growth and hyper urbanization in developing countries, and government initiatives for traffic management under smart cities models. With the increase in the deployment of smart transportation solutions among the smart cities, the traffic management market is expected to gain major traction during the forecast period[67].

Although there are many kinds of business processes in the field of traffic management with accordance to the purposes of stakeholders mentioned in previously, we focus on **business processes of traffic managers and operators** as major actors in the field below:

- Program plan and priorities for traffic management
- Budget and programming to support needs of the program
- Program alignment to agency and regional transportation mission, goals, and objectives
- Coordination among program planning to other key planning activities (such as a long-range plan)
- Process to integrate regional- and network-based operations strategies (such as active traffic management and integrated corridor management)
- Performance outcomes to influence program plan and programming

Note that *a specific actor can play **multiple roles** in the same or different business models* as shown in Figure 16. For example, a company as a specific actor can be a navigation provider and the traffic content provider in the same or in different business models. The relevant actors

[67] Goodvision - Report on the Size of Traffic Management market, https://medium.com/goodvision/report-on-the-size-of-traffic-management-market-317284d5189e

share their own data with each other for it to be collected, processed and analysed. In return, they receive information to create services that are required by their business models.

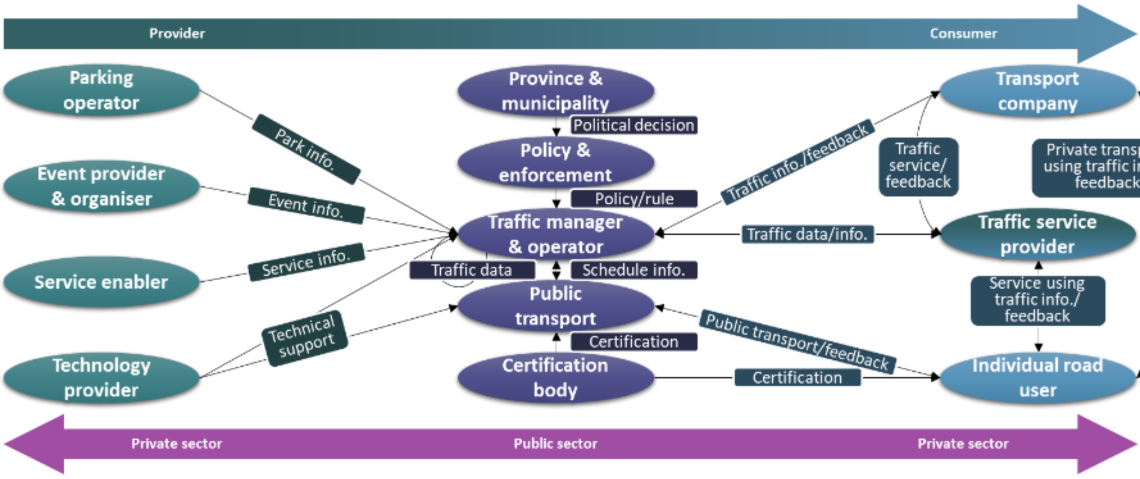

Figure 16 Business process in the ecosystem of traffic management

Business processes of traffic management are typically handled by one agency in a region, although in some areas there could be multiple traffic operators. However, roads represent a significant part of most regional transportation systems, so operations and management are most effective when they consider the road infrastructure with other modes and systems, such as transit and arterials. Regarding these, there might be potential challenges related to business process above, in terms of Big Data.

Table 14 Potential challenges for the business process of traffic managers and operators

Program elements	Potential business processes challenges of traffic management
Planning & programming	No short-term or long-range strategic plan introducing Big Data technologies exists for the traffic management program. Traffic management priorities are not developed collaboratively with key partners or with a network perspective for future leveraging Big Data. Budget planning for traffic management enhancements does not factor in new traffic data, technologies, strategies or approaches. Traffic management system life-cycles and asset management needs are not adequately addressed in agency budgets. Mainstreaming traffic management capabilities into capital programming (such as for traditional capacity enhancements or rehabilitation projects) might be not considered according to advance of technologies such as autonomous vehicles, AI and Big Data.

System operations	Traffic management processes have limited coordination with other partner operating processes (for example, law enforcement and arterial management agencies). Operations are largely static or by time of day and are not always able to respond to actual traffic conditions in real-time. Operations strategies are not scalable for big traffic data. Traffic management strategies do not incorporate decision-support tools or system performance for Big Data Analytics for traffic streaming data. Coordination among centres or systems in a region for traffic management is limited, particularly in the context of interoperability. Operational gaps are not addressed as part of the traffic management program that should specially consider characteristic change of traffic data, leading to business-as-usual operations approaches. Policy constraints limit implementation of more active traffic management strategies like adopting new technologies. Staff capabilities or skill sets are not aligned with usage requirement of advanced traffic management system such as ITSs, C-ITSs, autonomous parking systems and real-time TMSs.
Asset management	A preventive maintenance program is non-existent or is not adequate to address reliability of traffic management assets in terms of Big Data's veracity. Priorities compete within the agency for maintenance resources rather than considering advanced techniques to efficiently and effectively achieve the purpose. System maintenance, repair data, and device performance trends are not routinely used to inform budgeting processes, due to the lack of feedback from consumers. Maintenance and asset management responsibilities are distributed among multiple groups or divisions because of the absence of an integrated management system. Processes are not in place to accurately track and analyse device and equipment performance which could provide valuable inputs to budgeting and life-cycle planning.
Expansion & enhancement	Strategic planning for system expansion or future needs is limited. New strategies are not easily integrated into current road operations and management program, for example the combination of old DBMS and cloud server. Viable funding sources are not in place to address operational needs of system expansion (such as training for new strategies).

5.3.2 Relation to big data initiatives

Traffic management refers to measures to improve the flow and efficiency of dynamic traffic such as parking management, access controls, traffic guidance and signal control strategies. *Large-scale intelligent and interoperable traffic management systems are keys* to make better use of the capacity of the existing and future infrastructure and to optimise traffic flows with heterogeneous vehicles. In this regard, *there are many initiatives in the EU* when it comes to planning, technical solutions and development in transport sector[68].

In Sweden, Nicander is developing and providing ongoing support for Sweden's National Traffic Management System (NTS)[69]. NTS integrates urban, inter-urban and rural monitoring and control systems (over 60 systems and 10,000 devices of over 100 different types) such as smart motorways, tunnel and bridge control, roadworks schedules, automatic incident and congestion detection, vehicle access control, congestion charging, vehicle tracking and management, urban traffic signalling and signing, tidal flow, lighting, air quality, fire, asset and fault management and Supervisory Control and Data Acquisition (SCADA) into a single management platform providing advanced traffic management of the network. The system plays a role of overall network systems monitoring, analysing, making decisions and commanding systems and devices as a single management facility. In addition, the system integrates client databases for exchanging and managing information provided through DATEXII for third party service providers, radio data system-traffic message channel and in-vehicle systems.

The European Railway Traffic Management System (ERTMS) is an industrial project developed by eight UNIFE members (Alstom Transport, Ansaldo STS, AZD Praha, Bombardier Transportation, CAF, Mermec, Siemens Mobility and Thales) in close cooperation with the European Union, railway stakeholders and the GSM-R industry. ERTMS has two basic components: the European Train Control System (ETCS), which is an automatic train protection system (ATP) to replace the existing national ATP-systems, and GSM-R, a radio system for providing voice and data communication between the track and the train, based on standard GSM using frequencies specifically reserved for rail application with certain specific and advanced functions. ERTMS aims at replacing the different national train control and command systems in Europe. Each system is stand-alone and non-interoperable, and therefore requires extensive integration, engineering effort, raising total delivery costs for cross-border traffic. This restricts competition and hampers the competitiveness of the European rail sector vis-à-vis road transport by creating technical barriers to international journeys. Thus, ERTMS is designed to gradually replace the existing incompatible systems throughout Europe to bring considerable benefits to the railway sector as it will boost international freight and passenger transport.

In this regard, the collaborative decision making and system-wide information management proposed for air traffic management (SESAR), the advanced signalling and railway traffic management system (ERTMS), the safe and secure maritime traffic monitoring and information system (SafeSeaNet), the real-time river traffic information system (RIS) and cooperative intelligent transport systems (C-ITS), as well as initiatives on multimodal transport management and information systems plays a key role in speeding up the deployment of smart and intelligent mobility systems for improved traffic monitoring, control and communication to the traffic

[68] ITS - Action Plan and Directive, https://ec.europa.eu/transport/themes/its/road/action_plan_en

[69] National Traffic Management System (NTS) – Sweden, http://www.nicander.co.uk/wordpress/index.php/our-experience/nts/

controllers and vehicle operators. Research and innovation of new technological systems in this area is currently dealing with several parallel challenges:

- *European global navigation satellite system (Galileo)* has the potential to allow new opportunities for efficient tracing and tracking of vehicles such as pilots are currently being under investigation for train control. Consequently traffic management systems will have more frequent and reliable real-time information.
- *Digitalisation of information and improving data gathering* through monitoring-enabled components and actors is currently taking place. Dealing with Big Data also needs to be carefully managed and filtered, in order to be used effectively for real-time traffic management purposes.

5.3.3 Illustrative user stories

The utilization of Big Data driven real-time traffic management in transport sector varies largely dependent on the different transportation actors. The following stories are representative illustrations about generally encountered opinions:

- **Traffic service providers.** The main interest of owners is to provide competitive services using various techniques such as machine learning, AI and so on. However, most of them are only utilising basic techniques such as Dijkstra and A* algorithm, despite new data types are being generated in different time level. Therefore, in terms of real-time processing their services do not consider the real traffic situations in timely fashion.
- **Administrations/Authorities.** Transport authorities are mainly interested in aggregated data from vehicles and road environment. Even though they are finding companies and service providers able to provide advanced technologies for advanced traffic management, it is difficult due to gaps between budgetary restriction and high requirements.
- **Relevant events/facilities operators.** Although their activities (event) and functions (available parking space) have effects on traffic management, the data is often neglected due to the diversity of the information and the absence of real-time processing.

5.4 ANALYSIS OF SOCIETAL ASPECTS

5.4.1 Economic and political

According to INRIX's 2017 Traffic Scorecard report[70], the estimated total economic costs by traffic congestion for UK, Germany and the USA amounted to almost $461 billion in 2017. Since several major developing countries were excluded from the study, total global economic impact from traffic congestion could be significantly higher than the economic costs estimated by

[70] INRIX 2017 Global Traffic Scorecard Infographic, http://inrix.com/resources/inrix-2017-global-traffic-scorecard/

INRIX. In addition, regarding Future Market Insights, the economic cost of traffic congestion coupled with growing urbanization is a big problem. It also means that the traffic management market is a significant one, with a value of $5.4 billion and a CAGR of 18.2%[71]. ***The traffic management systems market is also increasing*** *due to increasing government focus on boosting safety as well as expanding the smart city initiatives along with rapid advancements in technology*.

With the growing traffic congestion levels across cities, the implementation of traffic management system is increasing exponentially. It has induced to develop advanced traffic management system devices as well as software to provide extraordinary capabilities and highly improved performances. For instance, the current technologies in traffic management system include integrated machine learning, wireless charging sensors, integrated toll management systems, weather monitoring solutions, IoT based ITS, ITS for connected vehicles, IoT for autonomous vehicles, and many more. *Owing to the speedy urbanisation and the numerous benefits of these various technologies,* **governments across various countries in the world are actively engaging themselves in the deployment of smart traffic management systems** *managing the internal road, rail, marine, and air traffic efficiently.* This is being done in order to smoothen traffic flow by reducing traffic congestion and reducing pollution levels across cities, by prioritising traffic in accordance with real-time traffic information.

Robust advancements in the IT industry *have created a strong growth potential for all the other industry.* At present, IT is entering all industry verticals with the latest technologies such as IoT, connected devices, machine learning, deep packet assessment, augmented & virtual reality and Big Data analytics. In particular, as various and huge traffic data are being generated with diverse speeds from these technologies, integrated Big Data solutions as traffic management systems will lead to the development of advanced systems with far more capabilities and enhanced performance, drastically enhancing the safety of individuals. Furthermore, market players now provide traffic management software integrated with machine learning and deep packet examination technologies, and have witnessed heavy demand for these software, thereby leading to healthy adoption.

However, *despite all the benefits from traffic management systems, the restricted budgets remain a serious issue*. Local authorities are often budget-constrained and try to save costs by avoiding the adoption of advanced traffic management systems. Furthermore, the authorities of some developing and underdeveloped regions do not have sufficient budgets to implement the systems[72]. In addition to these, the additional costs involved in installation, repairs, and maintenance of these systems require more expenses, eventually deteriorating the situation. For instance, because traffic management system requires multiple technology layers, municipal governments often lack the expertise in identifying and selecting the right mix of solutions. Thus, it would be usually essential to hire a project development contractor who is

---

[71] Future Market Insights - Traffic Management Systems Market: Deployment in Maritime Industry to Intensify in the Long Run Owing to Increasing Marine Traffic: Global Industry Analysis (2013 - 2017) & Opportunity Assessment (2018 - 2028),https://www.futuremarketinsights.com/reports/traffic-management-systems-market

[72] Future Market Insights - Smart City Initiatives to Augur Well for the Global Traffic Management Systems Market, https://www.futuremarketinsights.com/press-release/traffic-management-systems-market

an expert in designing and implementing traffic systems. *Consequently, **budgetary restrictions and high costs involved with advanced traffic management systems might be encumbering the growth of the market***.

5.4.2 Social and Ethical

Safety is a hard constraint to be satisfied by any network and traffic management system in society. There are multiple development directions including guidelines, directives and regulations discussed in next section to prioritise safety aspect by adopting, refining and incorporating systems designed to prevent accidents and collisions. For instance, in road traffic management, this can be observed in autonomous/semi-autonomous vehicle sensors, intelligent speed adaptation systems, etc. In rail traffic management, safety is even more important in the traffic management systems via automatic train protection systems, advanced trajectory planning and modern signalling systems. Although there are the concerns of people about safety of autonomous systems, traffic management systems could be means enabling an efficient monitoring, repairing and maintenance of the traffic data, infrastructure and vehicles. *However, technological challenges aside**, strict regulatory policy and reliability testing requirements** in changing a city's infrastructure can impede the deployment of new technologies.* As aforementioned, municipal governments also have limited budget for major radical infrastructure upgrades. They are also more conservative than the private sector, with city officials often more resistant to change and adopting new technologies. Finally, government procurement procedures often require success case studies, and it can be translated to a chicken vs. egg issue in terms of adopting innovative technologies such as Big Data, IoT, etc.

On the other hand, drone as one of technologies gathering more and more attention and being used in various fields (e.g., environmental monitoring, traffic management, pollution monitoring and civil security control) is rising as a critical issue in air traffic management systems[73]. For example, as commercial drones take flight more and more across Europe, regulators are working to establish standardized drone rules and traffic management best practices across the Europe continent. Although using drones can provide many benefits to smart traffic management, there are many safety, security, privacy, and ethical issues involved on these applications[74][75]. For instance, regarding the right of human life, everyone can be gotten monitored, photographed, tracked and targeted[76]. Furthermore, by the advances of the autonomous vehicle technologies having advanced sensors along with the wireless communication allowing vehicles to share information with other vehicles/infrastructure, there is huge expectation that the vehicles can cooperate each other and manage themselves through

---

[73] Air Traffic Management, https://airtrafficmanagement.keypublishing.com/magazine/view-issue/?issueID=7671

[74] Finn, Rachel L., and David Wright. "Unmanned aircraft systems: Surveillance, ethics and privacy in civil applications." Computer Law & Security Review 28.2 (2012): 184-194.

[75] Vattapparamban, Edwin, et al. "Drones for smart cities: Issues in cybersecurity, privacy, and public safety." 2016 International Wireless Communications and Mobile Computing Conference (IWCMC). IEEE, 2016.

[76] Whitehead, op. cit., 2010.

smart traffic management[77][78]. The technological developments are forcing government and society to reflect on the emerging changes in ethical aspect. For instance, the licensing for automated driving systems, the dependency on autonomous driving, the transparency and autonomy of traffic data to be used[79]. Like these, *since traffic management systems are closely related to other* ***state-of-the-art technologies*** *that could lead many ethical issues, ethical aspects should be also considered for utilising the traffic data generated from such techniques*.

The table below list the societal and ethical issues in businesses of TDT and Entur.

Table 15 Societal and ethical issues in businesses of TDT and Entur

Aspect	Description of aspect	Relevance of issue to TDT and Entur
Trust	Increase of decrease in trust in by or between the data users, data facilitators and data suppliers.	Supporting to gather correct feedback about traffic is important to increase trust of traffic information and it relies on fund from government. Both authorities are cooperating with other service providers to deal with this issue. In addition, autonomous traffic management systems (e.g., road sensor network) could help to reduce budget required.
Surveillance	The ability to collect large amounts of data enables the data user to better understand processes in society.	Although collecting huge data could be used in right way, utilising and tracking these data should be done with thorough validation for the data. Regarding two interviews with TDT and Entur, it is difficult to completely confirm data flow currently, especially in the situation involving vary stakeholders.
Privacy	The involvement of privacy sensitive information	Although various data such as vehicle ID might have personal information, data from CCTV is mostly exposed to the privacy issue. Therefore, both authorities have installed CCTV on the place which is essential to collect and manage traffic such as intersection, ticket machine, entrance of building and roads. It could reduce the risk of privacy and improve the safety capacity of society.

[77] El Hamdani, Sara, and Nabil Benamar. "Autonomous Traffic Management: Open Issues and New Directions." 2018 International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT). IEEE, 2018.

[78] European Commission - European Urban Mobility, https://ec.europa.eu/transport/sites/transport/files/2017-sustainable-urban-mobility-policy-context.pdf

[79] BMVI – Ethics commission automated and connected driving, 2017, https://www.bmvi.de/SharedDocs/EN/PressRelease/2017/084-ethic-commission-report-automated-driving.html

Free will	Big data analysis and associated predictive analysis that threatens human autonomy	Consideration of free will is necessary, when data is related to people or their mobility. In transport sector, low quality of traffic information could lead threatens of free will. For example, an improper decision of traffic restrictions based on analysis of traffic data will waste resources of time and come back with large amounts of claims. This situation expected by case study providers is one factor hindering to adopt advanced technologies of traffic management in real industry.
Personal data ownership	Confusion of who owns the data or loss of data ownership by data suppliers	Most of traffic data for traffic management is collected by infrastructure of competent authorities in real time. Therefore, no specific issues with personal data ownership.
Discrimination	Discrimination of population groups based on generalisations derived from big data	No issues related with discrimination were found.

5.4.3 Legal

Security and cyber-security issues *unavoidably arise*, especially in highly frequented public transport environments, such as highway car parks, major railway stations, truck and air hubs, busy inland ports and seaports. In Europe, excellent examples of highly secure environments are the metro systems of densely populated cities, such as the London Underground, or European airports. A similar level of security is envisaged in any intra-mode transport and in inter-modal connection points. Moreover, a high level of cyber-security is required to prevent hacking, jamming and unauthorised manipulation of traffic management and network operation systems, while also ensuring a satisfactory level of data security and privacy in transport, taking into account all, technological and legal constraints and requirements.

In the case of Entur, they maintain security protocols and policies such as the Payment Card Industry Data Security Standard to ensure cyber-security and they use cloud service providers which offer better security functions than local service providers usually do. There is also a team working on the maintenance and security of their platform, which they do by monitoring intrusion trails.

Ensuring **privacy of personal data** in traffic management systems is essential for all data subjects, transit agencies, governments and stakeholders involved. Since the data might contain personal information and could be used to track people and vehicles, one key challenge is to avoid fraudulent entities to generate and distribute fake warning messages, as was mentioned in the interviews with Entur and Transport Department in Tallinn. Data authentication or verification of data consistency, which checks the legality and consistency of messages to avoid messages with malicious data, could be one of solutions for these issues.

New safety rules enabling connected and automated mobility on EU roads traffic management have been adopted by the European Commission in 2019, as one part of the deployment of C-ITS[80]. The rules are in line with the EU's long-term goal of achieving close to zero fatalities and serious injuries by 2050 ("Vision Zero[81]"). The specifications from the EC establish the minimal legal requirements for interoperability between the different cooperative systems used such as autonomous driving systems and traffic management systems. Interoperability will enable all equipped stations to exchange data/messages with any other station securely in an open network. In addition, the cooperative element – enabled by digital connectivity between vehicles, and between vehicles and the transportation infrastructure/traffic management system – is expected to significantly improve road safety, traffic efficiency and comfort when driving, by helping the driver to make the right decisions and adapt to the traffic situation.

To analyse the legal issues, the framework from (Debussche, César, & De Moortel, 2018) is used as shown in *Table 16*. They identify the following legal dimensions of big data:

Table 16 Legal issues in businesses of TDT and Entur

Aspect	Description of aspect	Relevance of aspect to TDT and Entur
Privacy and data protection	Consideration and implementation of the requirements of the GDPR and other types of privacy and data protection-related legislation	Both authorities made significant efforts to process personal data in compliance with the applicable privacy legislation and are waiting the enactment of local legislation. .
(Cyber-)Security	Awareness and implementation of certain general or specific (cyber-)security requirements	They have a special team whose task it is to comply with rules and laws with regard to security and cyber-security. In the case of Entur, cloud servers including mature security functions are being used to store traffic data.
Anonymization and pseudonymisation	The use of any type of anonymisation or pseudonymisation technique, i.e. techniques to ensure personal data can no longer be attributed to individuals without using additional information or to reduce the identifiability of individuals	Although they in principle do not process any personal data, data which could identify a data subject are anonymised.
Supply of digital content and services	Uncertainties regarding supply of digital services for personal data	Since the quality of the traffic data might be low, it could be difficult to actively use such information for a relevant service.

[80] Traffic technology today.com, new safety rules on EU roads, https://www.traffictechnologytoday.com/news/legal/ new-safety-rules-enable-connected-and-automated-mobility-on-eu-roads.html

[81] Vision Zero Initiative 2050, https://trimis.ec.europa.eu/?q=project/vision-zero-initiative#tab-outline

Free flow of data	Awareness of any data localisation requirements hindering the free flow of non-personal data for the organisation	No localisation requirements of data are applicable to both authorities, however some open data are still published in Norwegian and Estonian.
Intellectual property	Protection by the organisation of its service/product through intellectual property rights and hindrance for the organisation of third parties claiming or otherwise asserting intellectual property rights	Since relevant data for traffic management is controlled by public organisations, future claims of intellectual property are unlikely.
Open data	Awareness and use of open public sector information	Both authorities' traffic data are not considered to have sensitive information, hence its availability is not at risk.
Sharing obligations	Awareness of and compliance with certain specific data sharing obligations, mandatory legal requirement to share data with other market actors	As NAP of Norway and Tallinn transport, both authorities comply with relevant legislations from the EU, Norway and Tallinn. Especially, Tallinn has initially constructed traffic management with the help from systematic transport data structure of Norway.
Data ownership	Third parties claiming data ownership	No issues related to data ownership are part of traffic management by competent authorities.
Liability	Use of data from non-public organisations and the organisation of the data sharing	Transportation networks are closely linked to each other. Therefore intervention of authorities having power from governments such as TDT and Entur is essential to the effectiveness and efficiency of traffic management.
Competition	Consideration of legal liability framework and any particular issues in this respect	No issues related to competition are part of their businesses.

5.4.4 Environmental

Taking into account the Europe 2020 strategy, Europe needs to make better use of the available resources under emission-reduction commitments. The 2050 Energy Union policy[82] objectives and targets include the quantitative elements such as decarbonisation (-60% GHG emissions and -40% domestic GHG emissions), efficiency (27% energy savings) and jobs, growth & investment. The transport sector plays an important role in this context, since it accounts for about 1 out of 4 of all GHG emissions, about 1/3 of all energy consumption. *The 2011 Transport White Paper[83] acknowledged that the development of a competitive, intelligent, multimodal, integrated and resource efficient* ***transport system requires advanced traffic management capabilities****, in order to contribute considerably to the reduction in CO2 emissions and a comparable reduction in oil dependency by 2050.*

At present, there are numerous bottlenecks within the four modes of transport such as air, rail, water and road transport. And they lead to undesirable operating environments producing non-optimal levels of transport performance, severe capacity constraints, unmanageable traffic jams and environmental/emissions outcomes. For example, on urban roads, a significant improvement of mobility of people and transport of goods requires a better management of all kinds of vehicles (from conventional to autonomous vehicles), bicycle & vehicle sharing, road public transport & paratransit, and walking & cycling. This issue particularly become a difficult problem in the dense urban and metropolitan environments. Thus, *the importance of efficient traffic management is more increasing with growing urbanisation.*

As one example of the transport activities to contribute environment, Tallinn is participating in the CIVITAS MIMOSA project meeting the growing mobility demand in a sustainable fashion, to make half CO2 emissions from transport. The bulk of the emission savings, around 40 percent, will come from the introduction of new cleaner vehicles. Another 10 percent will come from an eco-driving programme. The eco-driving campaign also intends to lower accident rates by 10 percent and accidents at remodelled pedestrian crossings are projected to decrease by 25 percent. Tallinn is drawing on its MIMOSA partner Bologna’s experience with cutting down on the misuse of bus lanes. The introduction of a new video surveillance system is projected to reduce illegal use of bus lanes by 70 percent. The city expects a 7 percent increase in the share of public transport through the promotion of collective transport[84].

[82] 2050 Energy Union policy, https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union/2050-energy-strategy

[83] Transport White paper 2011, https://www.eea.europa.eu/policy-documents/transport-white-paper-2011

[84] MIMOSA project, https://civitas.eu/city/tallinn

5.5 ASPECTS RELATED TO OPPORTUNITIES, BARRIERS AND LIMITATIONS

5.5.1 Opportunities

Data is a major challenge for transport and network planners, including Big Data collection/fusion/management, floating vehicle data, data collection via social media, etc., as found by interviews with TDT and Entur. Increasing **real-time information availability** can create a seamless connection. Such information has to be verified, filtered, elaborated and communicated via customised interfaces to the user to avoid privacy issue.

Advance of the level of automation of vehicle-to-vehicle and vehicle-to-infrastructure connectivity plays a key role in automated traffic management systems, using common standards and technical specifications, for example driverless metro lines in Paris, Lille, Rennes, Lyon and Toulouse[85]. Moreover, with these automated environment, **decision support systems** have the potential to support the traffic controllers towards optimised solutions via the use of a sophisticated traffic flow model, allowing not only the high quality reproduction of current traffic and traveller conditions, but also optimising the future traffic situation. Both TDT and Entur seem to carefully consider such advanced technologies for their tasks related to traffic management in terms of technical, societal and legal aspects.

Several developed countries and smart cities are already using IoT and Big Data to their advantage to minimize issues related to traffic. *With following the superior cases*, better-organized traffic systems mean **a better flow** of vehicles on the road, and it means no idling cars, buses, and trucks in traffic jams. All this eventually translates **to lower run times, proper utilization of natural resources (gas), and less pollution**. The emittance of gases is the largest during stop-start driving, and this happens in spots where traffic is regulated by lights. Hence, if you go for smart traffic, this helps in pollution reduction throughout the entire city.

5.5.2 Barriers and limitations

Despite many initiatives and projects supporting as country (e.g., Norway and Estonia) and large county (e.g., Oslo and Tallinn), the **budgetary restrictions of local areas** for adopting and implementing real-time traffic management systems using Big Data technologies and **the lack of human resources** for operating the systems might hinder integration of regional distributed systems into a connected traffic management solution. In addition, when **real-time analytic** comes to transportation, current ITS infrastructure and platforms deployed are not designed for real-time data processing. Reducing response time in traffic management systems is critical to ensure that information and decision making are tightly coupled in time and thus, useful in practice. *Real-time processing is therefore deemed crucial to provide new forms of on-demand mobility requiring real-time data sharing and interpretation, calling for legislative/public attitude shifts to implement real-time traffic management.*

Security and privacy are long-standing issues widely studied within transportation and mobility domains. Notwithstanding this prior work, the advance of recent technologies (connected and

[85] Around the world: 1,000km of fully automated metros, https://www.railway-technology.com/features/around-world-driverless-metro-lines/

autonomous vehicles, IoT and so on) generating Big Data in such domains has made these two interrelated aspects become even more complicated to address effectively. Furthermore, trends such as the usage of payment data (in all their forms and scales) as an additional source from which to infer macroscopic mobility patterns increase significantly the privacy risks. These privacy issues jeopardise a wider adoption of Big Data technologies for traffic management solutions[86].

As aforementioned, **new sources of data** generated by state-of-the-art technologies closely related to traffic management unleash new possibilities, services and applications with various challenges in terms of Big Data such as volume and coverage, variety and quality:

- **Data volume and coverage**: the current transportation datasets are much larger than before. This augmented information generated by vehicles implies that traditional traffic management technologies and processes to deal with the data do not work in practice, so it becomes necessary to rethink and redesign advanced traffic management systems to accommodate such volumes of captured data.
- **Data variety** is the complexity of multiple data sources relating to the rising tide of different formats, different temporal resolutions and varying levels of accuracy in traffic management solutions. The collected data are analysed and fused so that the system can provide society with optimal traffic management based on the aggregated information.
- **Data quality** refers to the completeness, homogeneity and/or trustworthiness of the data. With many forms of big traffic data, quality and accuracy are less controllable, as large data volumes often make up for the lack of quality or accuracy. A challenge of paramount significance is to provide the **standard** means to quantify the reliability and quality of data at both collection and retrieval stages, as well as to trace and measure their impact in the knowledge inferred in upper layers of the processing stack.

5.6 SOURCES OF INFORMATION

5.6.1 Interview

Interviews were carried out in the Tallinn Transport Department with:

- Department head
- Chief specialist
- Development specialist

Interviews were carried out with the Professor of the Research Centre of Logistics and Transportation of the Tallinn University of Technology.

Two interviews with Entur A/S were carried out with

- Chief Technology Officer
- Data analyst

[86] Torre-Bastida, Ana Isabel, et al. "Big Data for transportation and mobility: recent advances, trends and challenges." IET Intelligent Transport Systems 12.8 (2018): 742-755.

5.6.2 References

Wences, Pedro, et al. "Decision-Making Intelligent System for Passenger of Urban Transports." International Conference on Ubiquitous Computing and Ambient Intelligence. Springer, Cham, 2017.

Guan, Shuqi, et al. "Intelligent transportation system contributions to the operating efficiency of Urban traffic." Journal of Intelligent & Fuzzy Systems 31.4 (2016): 2213-2220.

Wang, Zhenhua, Yangsen Yu, and Dangchen Ju. "Analysis and Prediction of Urban Traffic Congestion Based on Big Data." International Journal on Data Science and Technology 4.3 (2018): 100.

Goodvision - Report on the Size of Traffic Management market, https://medium.com/goodvision/report-on-the-size-of-traffic-management-market-317284d5189e

ITS - Action Plan and Directive, https://ec.europa.eu/transport/themes/its/road/action_plan_en

National Traffic Management System (NTS) – Sweden, http://www.nicander.co.uk/wordpress/index.php/our-experience/nts/

Commission Delegated Regulation (EU) No 886/2013, https://eur-lex.europa.eu/legal-ontent/EN/ALL/?uri=celex% 3A32013R0886

Commission Delegated Regulation (EU) No 962/2015, https://eur-lex.europa.eu/legal-ontent/EN/ALL/?uri=celex% 3A32015R0962

RDS-TMC: Radio Data System – Traffic Message Channel

EU EIP SA46 Annual NAP report – 2016, https://www.its-platform.eu/filedepot_download/1971/6019

Intelligent transport systems – National Access Point, https://ec.europa.eu/transport/themes/its/road/action_plan /nap_en

Trans-European Transport Network (TEN-T), A European Commission policy directed towards the implementation and development of a Europe-wide network of roads, railway lines, inland waterways, maritime shipping routes, ports, airports and rail-road terminals. It consists of two planning layers: comprehensive and core networks, https://ec.europa.eu/transport/themes/infrastructure/about-ten-t_en

Commission Delegated Regulation (EU) 2015/962, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32 015R0962

National Access Point (NAP), https://ec.europa.eu/transport/themes/its/road/action_plan/nap_en

Wikipedia – NeTEx, https://en.wikipedia.org/wiki/NeTEx

Wikipedia – CEN, https://en.wikipedia.org/wiki/European_Committee_for_Standardi zation

EU regulatory and use of Transmodel, http://www.transmodel-cen.eu/overview/use-of-the-transmodel/

Transport information of Tallinn, https://www.tallinn.ee/eng/otsing?sona=11964

ITS national reports, https://ec.europa.eu/transport/themes/its/road/action_plan/its_national_reports_en updating the Working Programme in relation to the actions under Article 6(3) of Directive 2010/40/EU, 2018, https://ec.europa.eu/transport/sites/transport/files/legislation/c20188264_en.pdf

CEF transport projects by country can be find by https://ec.europa.eu/inea/connecting-europe-facility/cef-transport/projects-by-country

TURBLOG. (2011). FP7 TURBLOG Deliverable 2 Business Concepts and Models for urban logistics.

INRIX 2017 Global Traffic Scorecard Infographic, http://inrix.com/resources/inrix-2017-global-traffic-scorecard/

Future Market Insights - Traffic Management Systems Market: Deployment in Maritime Industry to Intensify in the Long Run Owing to Increasing Marine Traffic: Global Industry Analysis (2013 - 2017) & Opportunity Assessment (2018 - 2028),https://www.futuremarketinsights.com/reports/traffic-management-systems-market

Future Market Insights - Smart City Initiatives to Augur Well for the Global Traffic Management Systems Market, https://www.futuremarketinsights.com/press-release/traffic-management-systems-market

Air Traffic Management, https://airtrafficmanagement.keypublishing.com/magazine/view-issue/?issueID=7671

Finn, Rachel L., and David Wright. "Unmanned aircraft systems: Surveillance, ethics and privacy in civil applications." Computer Law & Security Review 28.2 (2012): 184-194.

Vattapparamban, Edwin, et al. "Drones for smart cities: Issues in cybersecurity, privacy, and public safety." 2016 International Wireless Communications and Mobile Computing Conference (IWCMC). IEEE, 2016.

Whitehead, op. cit., 2010.

El Hamdani, Sara, and Nabil Benamar. "Autonomous Traffic Management: Open Issues and New Directions." 2018 International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT). IEEE, 2018.

European Commission - European Urban Mobility, https://ec.europa.eu/transport/sites/transport/files/2017-sustainable-urban-mobility-policy-context.pdf

BMVI – Ethics commission automated and connected driving, 2017, https://www.bmvi.de/SharedDocs/EN/ PressRelease/2017/084-ethic-commission-report-automated-driving.html

Traffic technology today.com, new safety rules on EU roads, https://www.traffictechnologytoday.com/news/legal/ new-safety-rules-enable-connected-and-automated-mobility-on-eu-roads.html

Vision Zero Initiative 2050, https://trimis.ec.europa.eu/?q=project/vision-zero-initiative#tab-outline

2050 Energy Union policy, https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union/2050-energy-strategy

Transport White paper 2011, https://www.eea.europa.eu/policy-documents/transport-white-paper-2011

MIMOSA project, https://civitas.eu/city/tallinn

Around the world: 1,000km of fully automated metros, https://www.railway-technology.com/features/around-world-driverless-metro-lines/

Torre-Bastida, Ana Isabel, et al. "Big Data for transportation and mobility: recent advances, trends and challenges." IET Intelligent Transport Systems 12.8 (2018): 742-755.

6 Case study 4: "Logistics and consumer preferences"

Rapidly changing customer demands, behaviour and preferences make the end-to-end supply chain a focus for companies aiming at operational efficiency. Within the supply chain, the transportation component plays a critical role in ensuring that customer requirements and expectations are met. As a consequence, there are multiple challenges and demands for the commercial freight transportation and logistics industry.

In this view, this case study involves an American company Kepler51 which has been using advanced predictive analytics technologies to build a real-time logistics tool to increase the efficiency of delivery vehicles. Their big data solution - the LiveRoad Geospatial Analytics Platform - allows for the real-time monitoring and forecasting of risks and delays based on a range of factors (such as weather, temperature, road conditions, departure time, historical analysis, etc.), in order to dynamically route or schedule vehicles for efficient movements.

6.1 CASE STUDY OVERVIEW

6.1.1 Organisation/Stakeholders

Supply Chain Management (SCM) can be divided into three main areas: purchasing, manufacturing and transport. Transportation services are the essential component of SCM and cost-efficient logistics management can be a real point of competitive differentiation.

The following figure demonstrates the role of transport in supply chain management.

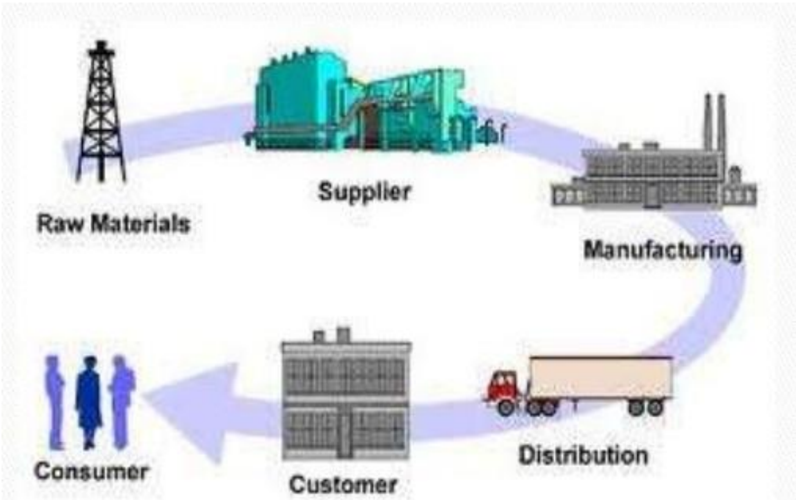

Figure 17 Role of transportation in supply chain management[87]

[87] Source: Transportation analysis in supply chain management: https://www.slideshare.net/7837686478/transportation-ppt-of-suppy-chain-management

Figure 18 Elements of the supply chain management[88]

About Kepler51

Kepler51 is a **big data analytics and meteorological technology company**. Their work combines meteorology with big data analytics. Kepler51 is a **public benefit corporation**: in the United States (US), they are a C-corporation and as such they can elect to have a public benefit. The specific public benefit of the C-corporation is to promote safer roads and increased public mobility through technology solutions designed to achieve fewer accidents, fewer injuries, and fewer fatalities.

They operate in both the **commercial** space and in the **passenger vehicle** space. Composed of a team of 15 people, Kepler51 is **headquartered in Austin, Texas**, but they also have their offices in the Bay Area and in Belgrade, Serbia. Their services are only deployed in the US however. Across the US, they have their staff distributed in different cities, for example in Detroit, working closely with the local OEMs (Original Equipment Manufacturer).

The Kepler51 team is composed of 4 meteorologists and 8 technical staff members, including data scientists in the front-end and back-end development. Other members include 3 people working in sales, marketing and administration, respectively.

[88] Source: https://www.infodiagram.com/diagrams/logistics-icons-transport-location-supply-chain-clipart-powerpoint.html#diagram-content

The beginning of the activities of Kepler51 in transportation and navigation can be traced back to 2006. The background of the two co-founders is in the development of weather alerts, which is how they noticed the limitations of using weather data alone. They understood that they can produce much more valuable information by combining data, removing the separation between different datasets and identifying the causal relationship between them, in order to understand what factors are increasing the risks on the road network.

In transportation, the key areas are vehicle safety and efficiency of both the vehicles and of transportation networks. They use machine learning and big data by looking at historical accident data on a broad scale across the US by bringing data from different states, running analysis against that accident data with historical weather data. They also run analysis against fleet delays, analysing when vehicles run late and how the weather has affected those drive times. Based on this, they develop predictive modelling that can predict increased risk based on the accident data as well as delays and disruptions on the network based on the forecasted conditions. They have been also working on a project related to fuel efficiency, both for gas-powered vehicles and electric vehicles. They analyse datasets from the customers or publicly available datasets and run very high-resolution weather modelling, using different machine learning techniques to develop the predictive models to predict increased risks, delays and fuel efficiency range.

Kepler51 applies their technology in the road transport sector, but also in the energy space, agriculture and retail. In transportation they work in fleet management and logistics, but they also have solutions moving towards autonomous vehicles, delivering the information into vehicles. They are currently in discussions to further develop this into advanced driver-assistance systems, such as adapted cruise control, i.e. adjusting the speed of the vehicle based on the expected environmental conditions or warning the vehicles that the vehicle sensors will not function in a certain area due to fog or some other event that will reduce the effectiveness of the sensors. In the energy sector, the predictor can be the amount of power that can be generated or outages on the network. In agriculture, the predictor can be yield, moss, mould or pest that can cause damage to the crops. In retail, an example of a predictor is the expected sales for a restaurant or retail store during different weather conditions.

6.1.2 Sector/Mode

This case study focuses on the road freight transport and logistics, but also incorporates several other industries that Kepler51 is works in, such as the retail, energy, and agriculture sectors.

6.1.3 Case study motto

The use of predictive analytics in fleet management for efficient logistics to meet customer preferences and requirements.

6.1.4 Executive summary

The **rapidly changing customer requirements and preferences** are creating new demand patterns for the commercial freight transportation and logistics industry. Shippers want logistics partners that can operate across their diverse supply chains and distribution networks and seek carriers that can maintain a high level of performance during disruptions. Shippers' supply chains are becoming ever more complex, even in market segments where their needs have been relatively straightforward in the past.[89]

The PwC 2017 report on the commercial transportation trends identifies a number of high-impact technologies that could address these challenges in the future, from self-driving trucks to robotics. **Real-time logistics** is one of such technologies: It will soon be possible to integrate trucks into logistics data across the entire supply chain. Advanced telematics will enable transportation companies, through cloud-based analytics, to track and monitor factors, such as truck location, the health and fatigue of the driver, the temperature and barometric pressure of the freight.[90]

This case study involves an American company Kepler51 which has been using advanced predictive analytics technologies in their real-time logistics solution. The case study focuses on one of their big data solutions - the **LiveRoad Geospatial Analytics Platform** which allows for real-time monitoring and forecasting of risks and delays based on a range of factors (such as weather, temperature, road conditions, departure time, and historical analysis), in order to dynamically route or schedule vehicles for efficient movements.

6.2 TECHNICAL PERSPECTIVE

6.2.1 Data sources/Uses

Kepler51 uses open datasets for their predictive analytics solutions. They use different types of **sensor data about accidents** but also sensor data that have to do with observation, such as **temperature**, **road conditions** and other similar observations. They also use **data from windshield wipers**[91] which is sent to them through APIs from their clients or from an OEM. This is private data which is not publicly available. The other **data** they use **is from road networks**, **observations on the road network** and a large collection of publicly available observation data. They also use the **nowcasting forecasts**, i.e. forecasts that are updated very frequently, on the order of every 15 minutes and that have a very high spatial resolution. They are take all the observations that exist from any source, including remote sensing from satellites, and combine it into a single view of what the road looks like every 15 minutes. In order to judge how valid the observations are, they look for outliers that do not make sense and then filter these out.

In the US, there is a long history of **data from road infrastructure stations**. Kepler51 is able to train statistical models that show the relationship between the general observation and the

---

[89] PwC report on Commercial Transportation Trends in 2016, cited in https://nowthatslogistics.com/adapting-to-changing-consumer-demands-with-a-flexible-supply-chain/

[90] Tipping, A. and Kletzel J. (2017) Commercial Transportation Trends: https://www.strategyand.pwc.com/trend/2017-commercial-transport-trends

[91] Windshield wipers triggering in mass in a particular area will be indicative of rain, butKepler51 also categorises this information in terms of a temperature in that area. At the same time, they have a certain amount of verification, for example, if one person turns on their windshield wipers it might not necessarily indicate rain, but just a person cleaning out their windshields. They correlate the events with other data they have in order to make a determination of what the meaning of this data is.

observation on the ground, such as the temperature for the last few days, the ground temperature in the area, and what the roads would be like in specific weather conditions (for example, after it has been raining for three days and then sunny for two hours and then icy later on). They have good models for making this type of predictions. Hence, it is not a major hindrance to their models and to provide good data to the customers, if they do not have access to data in a few States.

The US government sources provide **open datasets and data from the road network** that are required for innovative big data solutions. Kepler51 uses open data sources for their LiveRoad Geospatial Analytics Platform. They are able to develop their products relying solely on non-proprietary data, but they also use data provided by individual clients in their customer-based technologies.

Kepler51 does not utilise social media such as Twitter in its observations. However, they do use as a source the mPING (https://mping.nssl.noaa.gov/) network, a crowdsourcing app created by NOAA (the National Oceanic and Atmospheric Administration) where citizens can indicate the observations (e.g. the type of rain) they see through the window.

In terms of **machine learning**, their tool tries to establish a benchmark of normal behaviour and what a deviation of that norm is. This involves an estimation of how anomalous a certain road condition is, and based on that, determining potential hazards. For instance, in case of a two-week snow storm around the New Year's Eve in a certain area, they will look at the history of that occurrence over a period of 20 years. They will use this history to verify how anomalous it is that two inches of snow fell in a given area during that time of year and what the effects are on traffic and road conditions. Based on that analysis, they are able to estimate whether the infrastructure in this specific area is well placed to handle the conditions and whether the drivers have the awareness to handle these conditions. In a nutshell, they model these complex data using statistics and machine learning technologies.

In terms of historical records and **climate change affecting their predictions**, they have been utilising datasets from climate.org published by UC Berkeley. They have been analysing the climate using more recent information and not just relying on the full 30-year records. Further, they have recalibrated some of the climate information based on the work done by climate.org.

For their services developed for fleet customers, the real-time observations of drivers or data from the vehicle, may be taken into account when making predictions. The direction Kepler51 wants to go in is to have a live app where the information comes directly to them. In each use of private data sets, there are different restrictions, some more permissive than others, so data received from a certain client (company) is typically used to make predictions for that particular client only. Concretely, that means that they do not share customer data in hazards for other clients, unless agreed to by the customer for the benefit of all.

Table 17 Main data sources of the case study

Data source	Used in	Big Data dimensions	Other remarks
Traffic data from sensors on roads	road planning & design road maintenance & operation environmental tasks pollution from traffic (dust and noise)	Volume, Velocity, Veracity	Vehicle number, speed, light signal, traffic flow etc.
Weather data on roads		Volume, Velocity, Veracity	Weather, wind, temperature, snow in air and roadway.

Photos and videos from road CCTV cameras	landscape planning traffic safety work statistics about traffic accidents	Volume, Variety	From a part of CCTV tested for personal information.
National road data including toll, tunnel, road signals, etc.		Volume, not dynamic	Physical road network.
Traffic accidents		Dynamic, not big	Accident, vehicle, users levels etc.
Vehicle information		Dynamic, not big	Car types, etc.
Various reports		Not big and dynamic	Statistics data
Environmental data on road		Volume, Velocity, Veracity	Noises, emissions, etc.

6.2.2 Data flows

Data flows in logistics

From the point of view of a retailer, the aim is to have the product available in the shortest time possible with the correct labeling and supporting electronic documentation to allow those goods to seamlessly flow through their distribution network to their stores or customers. There are a number of challenges to this: suppliers, third-party transportation providers with varying levels of capability and performance, disconnected processes and technologies across the supply chain, etc.

The following figure demonstrates a simplified overview of data flows in logistics.

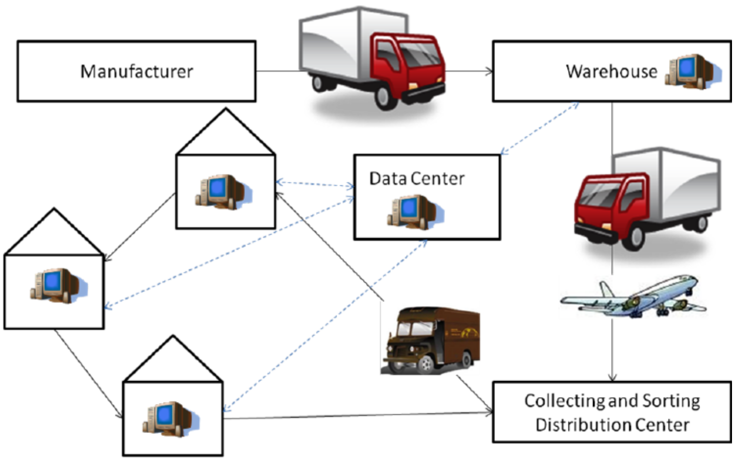

Figure 19 Data flows in logistics[92]

[92] Source: Weber C. L., Hendrickson C., Matthews H.S., Jaramillo P. Life cycle comparison of traditional retail and e-commerce logistics for electronic products: A case study of buy.com, Conference paper, 2009: https://www.researchgate.net/figure/E-commerce-Product-Flow-Diagram_fig2_224559288

Data flow in Kepler51's data centre

Data comes in Kepler51's data centre through their API network and then it goes to a number of different outlets. The data is normalised into internal proprietary formats using Google buffers, a format developed by Google for creating a structured binary dataset that is very compact. All the data that is pushed around their internal network goes into such binary data format. The data then gets pushed to Cassandra[93] database for permanent store and for analytics. The data is then put onto Spark, which is a real-time processing network, to be able to conduct anomaly detection and to categorise it in terms of significant events, and also to process the data in terms of hazards.

The hazards along the road network are indexed in terms of a geospatial index and then categorised in terms of the other events that have happened in the nearby area of the road network, along with the time around that event to determine whether the events are significant. Those are then processed into a data pyramid, which is a geospatial index that goes out in time. They have expanded the notion of a geospatial index to not only have current times, but also future times and over times. The resolution degrades: it starts at every 5 minutes and then it decreases as it goes forward, at 10 hours in the future for example there is a new index every hour. That means that as the time goes by, they decrease the resolution because the certainty of the events decreases. They are then able to make triggers on events that happened on the network. Once they have data in the spatial index, they are able to make predictions very fast, coming in from requests from the API.

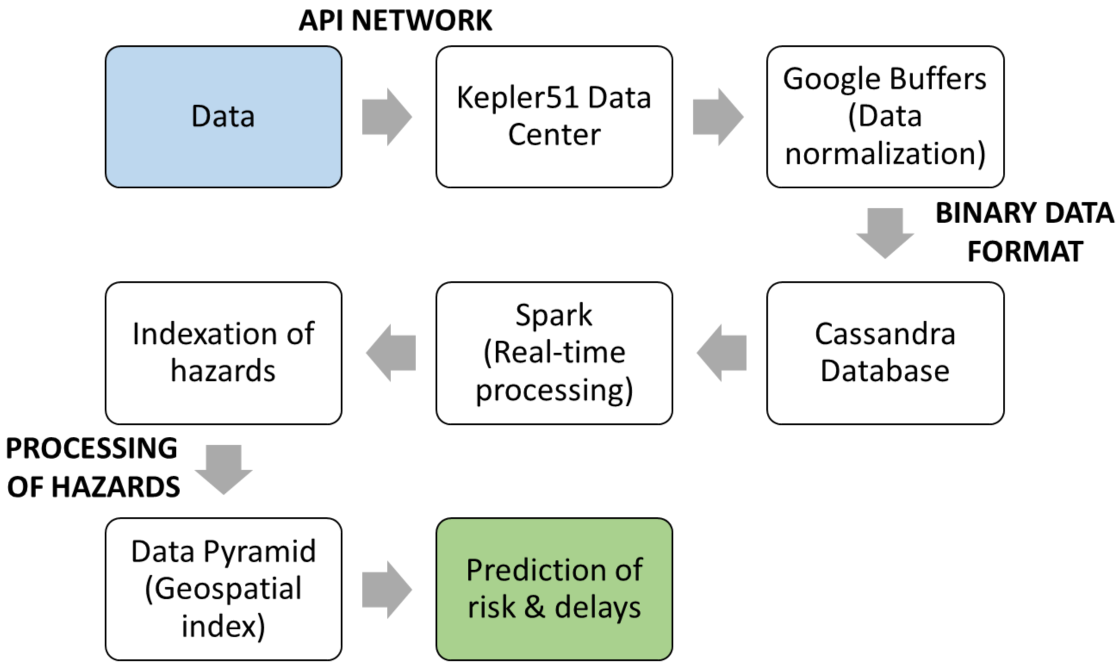

Figure 20 Data flow in Kepler51's data centre

[93] Cassandra is a free and open-source, distributed, wide column store, NoSQL database management system designed to handle large amounts of data across many commodity servers.

6.2.3 Relevant big data policies in the USA

Logistics and transportation industry in the US

The **logistics and transportation industry in the US**, the world's largest consumer market, is highly competitive. International and domestic companies in this industry benefit from a highly skilled workforce and relatively low costs. The US' **highly integrated supply chain** network links producers and consumers through **multiple transportation modes**, including air and express delivery services, freight rail, maritime transport, and truck transport. To serve customers efficiently, multinational and domestic firms provide tailored logistics and transportation solutions to ensure coordinated goods movement from origin to end user through each supply chain network segment.

The subsector of logistics services includes inbound and outbound transportation management, fleet management, warehousing, materials handling, logistics network design, inventory management, supply and demand planning, third-party logistics management, and other support services. Logistics services are involved at all levels in the planning and execution of the movement of goods.[94]

The **US freight transportation system has been facing challenges**, such as growing population, convoluted planning process and past underinvestment. To address these challenges, **U.S. Department of Transportation** (U.S. DOT) released a draft **National Freight Strategic Plan** in 2015, describing the physical, institutional, and financial barriers to improvement and outlining strategies to help support our freight transportation system through improved planning, dedicated funding streams, and innovative technologies.[95] The public was invited to comment until April of 2016. However, as of early 2019, there has been no further development of the Plan.[96]

Another document published by U.S. DOT - **the White Paper** from 2014 - concludes that **big data has significant potential for transportation operations**, especially in the transportation system monitoring and management where the utilisation of a wide variety of connected vehicle and connected traveler data to get the picture of real-time conditions enables much better predictions of impending conditions. One of the main conclusions of the document for the future is that understanding and capitalising on big data requires **broader, non-traditional and non-transportation stakeholder engagement**.[97]

When it comes to the **intermodal freight traffic**, it is increasing globally but policy settings and approaches differ from country to country. **In the EU**, the prevailing tendency is to try to force intermodal policies through a **top-down approach led by the European Commission**. In many

---

[94] The Logistics and Transportation Industry in the United States: https://www.selectusa.gov/logistics-and-transportation-industry-united-states

[95] U.S. Department of Transport. Our Freight Transportation System: https://www.transportation.gov/freight

[96] OnTrackNorthAmerica. Has Anyone Seen The National Freight Strategic Plan? https://www.ontracknorthamerica.org/2019/01/our-analysis-of-the-draft-national-freight-strategic-plan/

[97] U.S. Department of Transportation. Big Data's Implications for Transportation Operations: An Exploration, December 2014: https://rosap.ntl.bts.gov/view/dot/3542/dot_3542_DS1.pdf?

EU member countries intermodal transport is an important part and objective of sustainable transport policies often accompanied by modal shift actions diverting freight traffic from road to rail and, where feasible, to coastal shipping and waterways.

In NAFTA with the leadership of the **US**, on the contrary, **intermodal transport is driven by the market** and it is the business sector with large shippers and carriers, railway and shipping companies, forwarders and integrators, etc. that has pushed intermodal use without major governmental subsidies.[98]

Regulatory framework on big data in the US

There is **no single federal law** that regulates the collection and use of personal data in the US. Instead, federal regulation of big data is done through a variety of statutes. The government has approached privacy and security by **regulating only certain sectors** and types of sensitive information (e.g., health and financial), which often leads to overlapping and contradictory protections. Big data is affected by **state privacy laws** which can go further in privacy regulation than the federal government, especially privacy laws which directly address online disclosures and record keeping. One example of a robust state privacy law is California's Online Protection Privacy Act which requires websites to provide explicit privacy rights statements and allows users to know how their information will be used in the future. Other states have also begun to address privacy and data issues.[99] [100]

Companies operating in big data operations must ensure that their activities comply with privacy laws that are applicable to the data involved in their operations, as well as the companies' own privacy policies and all applicable contractual requirements. Companies also need to comply with all other applicable privacy and data security laws, including laws regarding privacy policy disclosures, laws concerning data breaches and regulations on data security requirements.

In order to minimize some of the risks associated with the use of big data, companies often decide to anonymize the data prior to analyzing them or sharing the information with third parties. Companies using data that is strictly anonymized still need to ensure that they comply with their own privacy policies and contractual obligations.[101]

6.2.4 Main technological challenges and issues

Transportation, and logistics in particular, are under constant pressure to remain competitive and deliver quality customer service, while finding innovative ways to optimise their operations.

[98] Horn, B.E. and Nemoto, T. (2005) Intermodal Logistics Policies in the EU, the U.S. and Japan: http://www.jterc.or.jp/kenkyusyo/product/tpsr/bn/pdf/no27-01.pdf

[99] O'Connor, N. Reforming the U.S. Approach to Data Protection and Privacy. Council on Foreign Relations, published on 30 January: https://www.cfr.org/report/reforming-us-approach-data-protection

[100] Myers, C. Big Data, Privacy, and the Law: How Legal Regulations May Affect PR Research, published on 11 June 2018: https://instituteforpr.org/big-data-privacy-and-the-law-how-legal-regulations-may-affect-pr-research/

[101] Klosek, J., Taylor Wessing Global Data Hub. Regulation of big data in the United States: https://globaldatahub.taylorwessing.com/article/regulation-of-big-data-in-the-united-states

Transportation costs in the company's overall logistics budget, challenges related to visibility, tracking, and management of the supply chain, the constant demand for faster and more efficient shipping strategies and increasing demand for segmented services and customer service are just a few examples of the many technological challenges that companies operating in the highly competitive logistics market are facing.

More specifically, in terms of their big data solution for fleet management, Kepler51 sees the main risk in terms of **confidence**, i.e. will the end-user be confident about the data. This is the main challenge of the new stage of weather prediction: it may happen that the representation of the world that the end-users actually see may not be synchronised with the ***nowcasting*** view that they provide. Therefore, the main challenge is to ensure that the end-users are confident about the model that is presented to them.

In terms of talent acquisition, Kepler51 is a relatively small company so they face some **challenges in terms of acquiring new talents** with the right set of mathematical, programming and meteorological skills, as the work that they do requires a more scientific subset of a typical programming/engineering profile that may be found in other technological companies.

6.2.5 Big data dimensions and assessment

There are a number of risks that come up when analysing the data. The data has a great deal of **errors** as the data input from different centres is highly **variable**. The data coming in from different data sources has a great variance in terms of the **certainty** of the timing of data, the **accuracy** of the centres and the accuracy of the source of the transmission. Different sources of data have a different precision and accuracy and it is a technical challenge to judge them and try to get a coherent vision of the world. By using data collected by third parties, there may be an amount of mismatch or collision so they need to look in detail in each case to identify what the problems are and to finetune the datasets from different data sources. In addition, data they use in their solutions also changes very quickly due to the nature of the weather so making the determination on **how long certain data is valid** influences the predictions.

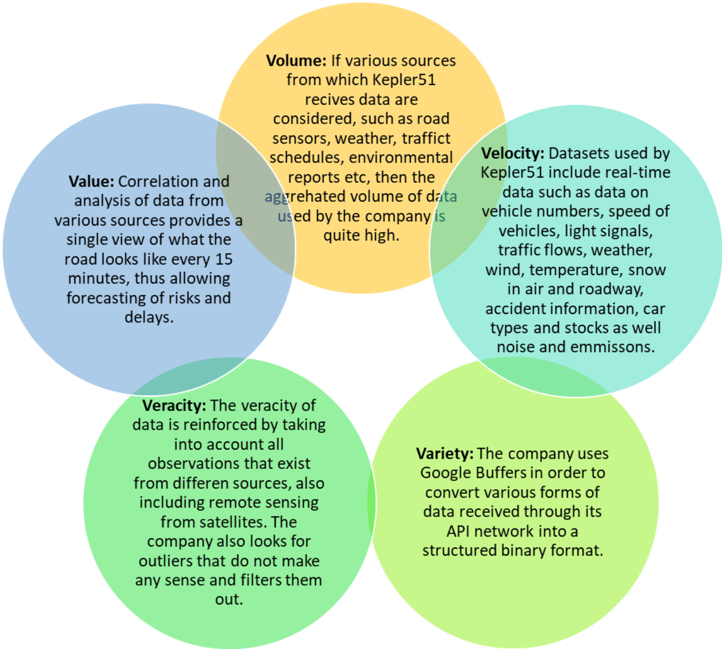

Figure 21 The V's of big data relevant for the case study

6.3 BUSINESS MODEL

Kepler51's business model is presented below.

Key partners	Key activities	Value Proposition	Customer relationships	Customer segments
US-government services that provide open datasets required for their technological solutions.	Building of tools and solutions matching specific needs of individual clients. Transposition of competences from one field to another and building of conceptual models. Sourcing of datasets. Data modelling. Collaboration with public authorities.	Big Data predictive analytics solutions (such as the LiveRoad Geospatial Analytics Platform) applied to road and transport to forecast risks and delays.	Kepler 51 uses a flexible approach with its customers, who pay per use for datasets and solutions provided by Kepler51 as per thier specific needs.	Companies in commercial and in the passenger vehicle space, agriculture, retail and energy that need tools to monitor disruptions on their networks to manage their operations.
	Key resources	**Internalization of costs and benefits**	**Channels**	
	Required datasets. Databases for storage of data and analytics. Processing network. Staff with the right set of mathematical, programming and meteorological skills.	No external costs.	Electronic format	
Cost structure			**Revenue streams**	
HR costs; costs for infrastructure used for storage and analysis of data; costs of infrastructure maintenance.			Pay per use.	

Figure 22 Business model canvas of Kepler51

6.3.1 Business processes

Kepler51 has a number of business models operating with different clients. They have observed that basically all actors in the market are looking for a business model for the use of their data, as they are trying to estimate the value of their data. Most of the actors are interested in the **pay-per-use**, meaning that as data is being consumed the data is being paid for, rather than in the past when there was a flat rate for particular clients. This opens up opportunities to have multiple types of datasets and multiple types of levels of solutions and to be able to offer solutions that can be more cost-effective or more expensive, depending on what level of data the customers want to go into a solution.

Having a **flexible business model** gives a lot of flexibility for clients to get started, but it also gives opportunity for other companies Kepler51 works with which can be paid as their data is being used. That gives the customers the freedom to decide what level of resolution they want, what level of accuracy they want in the product, what is their threshold (i.e. if they want to pay extra for e.g. extra 20% accuracy), etc.

6.3.2 Relation to big data initiatives

Demand for **information technology (IT) services in the logistics industry has increased** in recent years as more and more companies begin to deploy IT resources in their logistics/supply chain operations. The implementation of big data can lead to **safer supply chains** globally, with small to midsize importers and exporters being able to gain the most from new technology since they do not have the resources to create a safe supply chain but will be able to rely on freight companies that provide big data capabilities to customers. **Blockchain technology** can **increase transparency** for importers and exporters by replacing the need for extra time and personnel to record transactions within the supply chain and benefiting shippers to minimize information disruption, reduce paperwork needed, and maintain access to original information without worrying about tampering during and after shipment.[102]

In the EU, **the Communication on the Freight Transport Logistics Action Plan 2007** established a list of activities to improve the framework for transport logistics operations in the EU. Discussions with stakeholders identified that in the EU logistics costs represent about 10-15% of the final value of products. It is estimated that about half of these costs could be saved if obstacles were removed. These **obstacles** are in particular **high administrative burden** and **inefficient transport chains**, **lack of transport infrastructure** and the **non-completion of the internal transport market**.[103]

Digitalization, new technologies and big data have the potential to change the way cargo and traffic flows are organized and managed and enables cooperation between supply chain actors, better supply chain visibility, real-time management of traffic and cargo flows, simplification

[102] Sedat. Four Ways Logistics Will Change In 2019, published on 14 November 2018: https://www.forbes.com/sites/forbestechcouncil/2018/11/14/four-ways-logistics-will-change-in-2019/#2c6e42a43494

[103] European Commission. Logistics and multimodal transport - Logistics: https://ec.europa.eu/transport/themes/logistics-and-multimodal-transport/logistics_en

and the reduction of administrative burden, and allows for a better use of infrastructures and resources, thereby increasing efficiency and lowering costs.

To support this process, the European Commission (EC) established a dedicated expert group, the **Digital Transport and Logistics Forum (DTLF)**. The DTLF is a collaborative platform, where Member States, public entities and organisations exchange knowledge and coordinate policy and technical recommendations for the EC, in the fields of transport and logistics digitalization across all modes of transport.

The DTLF meets in Plenary as well as at the level of subgroups. The subgroups are set up to pursue specific issues, and are disbanded as soon as their mandate is fulfilled. During its first mandate (2015 – 2018), the DTLF prepared recommendations and carried out preparatory work for the proposal for a **regulation on electronic freight transport information**.

DTLF also prepared a concept of **digital corridor information systems** ('federative platform'), aimed at facilitating data sharing between all types of supply chain stakeholders by connecting existing cross-border IT platforms and services.

In September 2018, the EC adopted a Decision on setting up the second mandate of the Forum. The renewed Forum is expected to pursue its activities towards concrete implementation of the results achieved under the previous term, as well as to identify and prepare for other relevant actions in transport and logistics, aimed at creating EU added value and contributing to the completion of the Digital Single Market.[104][105]

6.3.3 Illustrative user stories

Kepler51 has been developing technical **solutions to monitor disruptions on the road network for various clients**. Their big data solutions - the LiveRoad Geospatial Analytics Platform - allows for comprehensive real-time monitoring of the road network and **forecasting of risks and delays** based on a range of factors (such as weather, temperature, road conditions, departure time, historical analysis, etc.

Based on the regression analysis on past late or missed service calls and contributing factors (analysing data such as day of week, time of day, atmospheric conditions, road conditions, etc.), their predictive model can identify potentially missed calls in the schedule and allows for the communication with the customer to be rescheduled, in order to efficiency of vehicle movements.

[104] European Commission. Logistics and multimodal transport - Digitalisation of Transport and Logistics and the Digital Transport and Logistics Forum: https://ec.europa.eu/transport/themes/logistics-and-multimodal-transport/digitalisation-transport-and-logistics-and-digital-transport-and_en

[105] Digital Transport and Logistics Forum. About us: http://www.dtlf.eu/about-us/presentation

6.4 ANALYSIS OF SOCIETAL ASPECTS

6.4.1 Economic and political

The economic value of technology developed by Kepler51 is derived primarily from the diminished disruptions on the road network and a more seamless flow of goods as a result of their fleet management solutions, and consequently reduced time and money losses in the system. This includes a reduced number of road accidents, reducing both their economic and human costs.

6.4.2 Social and ethical

Trust, privacy and transparency are some of the core values for Kepler51, and the company and its staff are committed to upholding these values in all their operations. In terms of **privacy and personal data ownership**, as an American company doing their technological development activities in Serbia, Kepler51 has observed certain **restrictions in Europe about using personalised data** when it comes to the data collected about cell phones. There is a great concern from the customers that are interested in understanding driver behaviour about restrictions, for example whether you can put an application into someone's phone or to track driving behaviour to be able to identify harsh acceleration, harsh bracing, stopping for stop signs, etc.

The table below lists the societal and ethical issues in business of Kepler51 according to the six topics identified by (Debussche, César, Hong, Nordeng, & Waldenfels, 2018): trust, surveillance, privacy, free will, personal data ownership, discrimination and environment.

Table 18 Societal and ethical issues in business of Kepler51

Aspect	Description of aspect	Relevance of issue to Kepler51
Trust	Increase of decrease in trust in by or between the data users, data facilitators and data suppliers.	Kepler51 sees one of the main risks in terms of confidence, i.e. will the end-user be confident about the data. The main challenge is to ensure that the end-user is confident about the model that they are producing.
Surveillance	The ability to collect large amounts of data enables the data user to better understand processes in society.	Kepler51 would wish to have access to more metrics from driver behaviour from tracking applications in mobile phones, but they are restricted on what information they can gather. There could be an overall benefit to society identifying unsafe drivers. However, this opens up numerous questions related to surveillance.
Privacy	The involvement of privacy sensitive information	Kepler51 relies mainly on non-proprietary, publicly available datasets and therefore do not face specific privacy-related issues.

Free will	Big data analysis and associated predictive analysis that threatens human autonomy	No perceived threat to free will was identified.
Personal data ownership	Confusion of who owns the data or loss of data ownership by data suppliers	Kepler51 relies heavily on open data, but also uses privately owned data provided by their clients in specific products for these particular clients. In such cases data ownership should be considered to avoid any confusion of owning original data.
Discrimination	Discrimination of population groups based on generalisations derived from big data	No issues related with discrimination were identified.

6.4.3 Legal

As a preliminary matter, it should be stressed that Kepler51 is a **US public benefit corporation** based in Texas which does **not deploy its product in the EU**. In light hereof, the legal framework applicable to Kepler51 differs substantially from the legal framework that has been identified in Deliverable 2.2, which only deals with the legal aspects of big data under the EU legal framework.

Kepler51 is heavily dependent on **open data**. The datasets they use typically stem from the US government and other public bodies. Overall, they have little difficulty in obtaining this data from the US public sector. However, one barrier that was identified is the fact that **different US states may open up different types of datasets**, leading to the fact that Kepler51 is able to access a certain type of data for a number of states, but not for all states. While overall, many datasets have been opened up; there is still some difficulty to obtain access where datasets remain closed for now. Having access to high quality data is crucial for the quality of the output, but Kepler51 is able to generate its product with the open datasets.

A key element for Kepler51 is to have **more real-time data**, typically **vehicle data**, from OEMs. This data is **privately held by the OEMs**, and Kepler51 is usually only granted access under strict conditions. In the same manner as in the EU, data sharing agreements provide the only solution to govern access to and/or exchange of data between the numerous stakeholders active in the big data value cycle. This can present issues, as one of the conditions that is often contractually imposed by the data-sharing company is the siloing of data. This hinders optimal efficiency and use of the big data analytics platform. This proves that the actual terms of a DSA may be so restrictive that the data recipient cannot optimally use the data as desired.

Additionally, it should be noted that this confirms the growing importance of real-time data, which had been identified as an issue by the EU legislator, and has been taken into account in the EU Commission's proposal for a new directive on the re-use of public sector information. That proposal includes the requirement for public sector bodies to make such data available through an Application Programming Interface (API), an obligation that would clearly also benefit Kepler51.

Kepler51 is however not always able to obtain access and companies asserting data ownership is the main barrier in this respect. Other barriers are usually more financial than legal.

The data used and processed by Kepler51 are mainly weather data, traffic data and other circumstantial data and does not constitute personal data. Consequently, they have not dealt with any legal requirements on privacy and personal data protection and have not implemented anonymisation or pseudonymisation techniques.

It follows clearly from the interview with Kepler51, which relies so heavily on open data, that the opening up of data has enabled them to access to resources which they would otherwise not be able to access. This has eliminated a barrier to market entry which would have been harder to overcome without such access to open data. The open data has allowed a company like Kepler51 to innovate and bring to the market a big data platform which helps increase road safety and eliminate societal costs of road accidents.

Overall however, legal requirements seem to be less strict in the US than in the EU.

To analyse the legal issues, the framework from (Debussche, César, & De Moortel, 2018) is used as shown in the table below. They identify among others the following legal dimensions of big data:

Table 19 Overview of legal aspects and relevance to Kepler51

Aspect	Description of aspect	Relevance of aspect to Kepler51
(Cyber-)security	Awareness and implementation of certain general or specific (cyber-) security requirements	As regards cybersecurity requirements, Kepler51 is not aware of any such requirements imposed through the applicable (US) legal and regulatory framework; Security requirements are mainly contractually imposed by customers wanting to protect their data.
Intellectual property	Protection by the organisation of its service/product through intellectual property rights and hindrance for the organisation of third parties claiming or otherwise asserting intellectual property rights	The company has no (registered) IP at the moment; There is a perception with Kepler51 that IP laws in the US are somewhat lighter than in the EU.
Open data	Awareness and use of open public sector information	Kepler51 relies heavily on open data; the opening up of data has enabled them to access to resources which they would otherwise not be able to access; Open data has therefore eliminated a barrier to market entry which would have been harder to overcome without access to such open data; This in turn has allowed the small company to innovate and bring to the market a big data platform which helps increase road safety and eliminate societal costs of road accidents.

Data ownership	Third parties claiming data ownership	Kepler51 is however not always able to obtain access and companies asserting data ownership is the main barrier in this respect.
Data sharing agreement	Use of data from non-public organisations and the organisation of the data sharing	Kepler51 often also uses privately held data (typically vehicle data from OEMs), in which case Kepler51 is typically only granted access under strict conditions; In the same manner as in the EU, data sharing agreements provide the only solution to govern access to and/or exchange of data between the numerous stakeholders active in the big data value cycle; Kepler51 has encountered in this respect, as one of the conditions that is often contractually imposed by the data-sharing company is the siloing of data, thereby hindering optimal efficiency and use of the big data analytics platform; This shows that the actual terms of a DSA may be so restrictive that the data recipient cannot optimally use the data as desired; Kepler51 itself does not share or disclose any of the data that it uses to feed its big data analytics.

6.4.4 Environmental

No particular environmental issues have been were found in the case study. Kepler51 is committed to respecting the environment and has been developing a solution as part of a project related to fuel efficiency, both for gas vehicles and for the electric vehicles.

6.5 ASPECTS RELATED TO OPPORTUNITIES, BARRIERS AND LIMITATIONS

6.5.1 Opportunities

As a public benefit corporation, Kepler51 collaborates closely with a number of different public authorities with the aim to improve road safety, including ICS America, the US Department of Transportation, highway administration and to a certain extent with the Commercial Vehicle Safety Alliance (CVSA). In terms of future opportunities, in addition to the work on road safety, they have also observed a lot of **Smart City initiatives** and a closer cooperation with municipalities and departments of transportation and they are looking to become a part of this process. All in all, it takes the **collaboration of the public and private industry partners** and with traffic and weather now becoming so critical with high-definition mapping, they can see a closer relationship between data providers and big data analytics solutions providers.

6.5.2 Barriers and limitations

For Kepler51, the key thing to overcome a number of limitations would be to get **more real-time data coming from the OEMs**. In the future they would like to establish a closer relationship with the companies that would allow us to use real-time information. Both in terms of prediction and in terms of nowcasting, the most important thing is getting **live sensor data from the road network** to be able to establish the baseline and then to utilise that to make short-term predictions and to correct other data sources that are coming in. An example consists of **infrared sensors from the road** that some manufacturers have, measuring the temperature of the road. Another example would be the **vehicle data** (information on the window wipers, fog lights, etc.).

They have observed a certain level of encouragement from the state towards to the OEMs to share more data in the context of the Smart Cities that companies like Kepler51 could profit from for their predictions. There is a consensus that this is something that needs to be addressed and the OEMs are coming into partnership, working closely with the public sector to help to realise that, so they are definitely seeing a movement of a closer collaboration on this front.

The companies would be interested to see information containing **more metrics on driver behaviour from tracking applications in mobile phones**. However, they are restricted on what information they can gather. There could be an overall benefit to society by being able to identify unsafe drivers. From the corporate side, they are not looking to identify and punish individual drivers, but rather to understand the causal factors and provide training for some sort of incentive programs. It is interesting that corporations may have a goal to be safer because they are paying the money because of the accidents, but they are restricted on what end-user data can they use. In words of the Kepler51 CEO, it all comes down to a question of how much personal data you can use to bring benefit to the society.

6.6 SOURCES OF INFORMATION

6.6.1 Interview

Two interviews with Kepler51 were carried out with

- Co-Founder & CEO
- Co-Founder & CTO
- VP of Automotive Solutions.

6.6.2 References

Debussche, J., César, J., & De Moortel, I. (2018). Report on legal issues.

Debussche, J., César, J., Hong, M., Nordeng, Z., & Waldenfels, R. (2018). Report on ethical and social issues.

Transportation analysis in supply chain management: https://www.slideshare.net/7837686478/transportation-ppt-of-suppy-chain-management

PwC report on Commercial Transportation Trends in 2016, cited in https://nowthatslogistics.com/adapting-to-changing-consumer-demands-with-a-flexible-supply-chain/

Tipping, A. and Kletzel J. (2017) Commercial Transportation Trends: https://www.strategyand.pwc.com/trend/2017-commercial-transport-trends

The Logistics and Transportation Industry in the United States: https://www.selectusa.gov/logistics-and-transportation-industry-united-states

U.S. Department of Transport. Our Freight Transportation System: https://www.transportation.gov/freight

OnTrackNorthAmerica. Has Anyone Seen The National Freight Strategic Plan? https://www.ontracknorthamerica.org/2019/01/our-analysis-of-the-draft-national-freight-strategic-plan/

U.S. Department of Transportation. Big Data's Implications for Transportation Operations: An Exploration, December 2014: https://rosap.ntl.bts.gov/view/dot/3542/dot_3542_DS1.pdf?

Horn, B.E. and Nemoto, T. (2005) Intermodal Logistics Policies in the EU, the U.S. and Japan: http://www.jterc.or.jp/kenkyusyo/product/tpsr/bn/pdf/no27-01.pdf

O'Connor, N. Reforming the U.S. Approach to Data Protection and Privacy. Council on Foreign Relations, published on 30 January: https://www.cfr.org/report/reforming-us-approach-data-protection

Myers, C. Big Data, Privacy, and the Law: How Legal Regulations May Affect PR Research, published on 11 June 2018: https://instituteforpr.org/big-data-privacy-and-the-law-how-legal-regulations-may-affect-pr-research/

Klosek, J., Taylor Wessing Global Data Hub. Regulation of big data in the United States: https://globaldatahub.taylorwessing.com/article/regulation-of-big-data-in-the-united-states

7 Case study 5: "Smart inland shipping"

It is not coincidental that an inland waterway-related ITS-based traffic management solution emerged in the Netherlands. The inland waterway (IWW) network in the Netherlands is the most dense and most intensive navigated network in Europe. In 2017, barges constitute almost 30% percent of all freight traffic in the Netherlands compared to circa 65% for road transport (Central Bureau for Statistics, 2019). The network performance of the IWW network and the road network are strongly interdependent. For instance, in the Netherlands, many movable bridges are located on its main highway network. Bridges form a major bottleneck for barges and road traffic. To avoid congestion in peak hours, bridges currently often do not open during peak hours. This limits the flexibility of inland waterway transport. When bridges are open for barges to pass, road traffic needs to wait before the bridge closes again, resulting in congestion on the road.

The case study on "Smart inland shipping" focused on the use of big data in a data platform called the Blauwe Golf Verbindend (BGV) established by three regional governmental organisations, the provinces of North-Holland and South Holland and the port authority of Rotterdam, and *Rijkswaterstaat*, the executive agency of Dutch Ministry of Transport. BGV aims to optimize the situation for both IWW vessel and road vehicles by optimizing IWW traffic through traffic management and providing IWW services. The platform supports the sharing of data between the road and inland waterway transport management systems. Interviews were carried out with Rijkswaterstaat (the transport ministry of the Netherlands), the province of North Holland, a shipping association, and a vessel captain. The results will help other data-driven initiatives in traffic management, some of which are specific to inland waterway management, while others apply to any type of traffic management.

7.1 CASE STUDY OVERVIEW

7.1.1 Organisation/Stakeholders

Traffic managers aim to ensure a good quality transport system that satisfies transport network demand (i.e. traffic) while talking into account the transport network supply (i.e. infrastructure). To achieve this, one of the methods available is real-time traffic management. The ability to influence in real-time specific traffic situations minimizes congestion and other negative externalities, and provides a better match between transport network supply and demand. In road systems, real-time traffic management is achieved by controlling traffic signals at intersections or ramps, dynamic highway lane and speed signs, detour messages, road pricing and prioritizing public transport or emergency services.

Traffic management systems (TMS) apply a top-down perspective to ensure traffic flow and road safety by using technology. Data such as data coming from speed sensors, lasers and cameras is integrated and processed into a central TMS. Based on this data, traffic-related decisions are made by (a combination of) action of traffic managers and automated systems. In response to the increase in transport demand and the pressure on resources for real-time traffic management, traffic managers started to design more intelligent transport systems. Data-driven technologies have come up as a solution to manage traffic more efficiently (Zhang et al., 2011). One of the most important new opportunities for traffic management that data offers is the continuous monitoring of vehicle mobility (Milne & Watling, 2018). Well known tools, such as GPS tracking systems, provide real-time insights on person- and vehicle movement through all times of the day. Having access to real-time traffic data enables traffic managers to adjust the situation to current traffic demand. While data-driven traffic management solutions are already found for road (De Souza et al., 2017), rail (Ghofrani, He, Goverde, & Liu, 2018), maritime (Zaman, Pazouki, Norman, Younessi, & Coleman, 2017) and air traffic control (Belcastro, Marozzo, Talia, & Trunfio, 2016), examples for data-driven inland waterway (IWW) traffic management are rare. This paper presents one such example established in the Netherlands.

In 2011 the Dutch ministry of transport launched a national grant programme called '*Beter Benutten"* (in English: *'Making better use of'*) to accelerate the deployment of intelligent transport solutions. One of the projects receiving a grant under this programme is the 'Blauwe Golf Verbindend ' (in English: '*Blue Wave Connected',* hereinafter referred to as BGV) project, which was established by three regional governmental organisations, the provinces of North-Holland and South Holland and the port authority of Rotterdam, and *Rijkswaterstaat*, the executive agency of Dutch Ministry of Transport. The reason for the establishment was the 50% co-financing received by the parties through the Beter Benutten program. The participating parties received subsidy for the realisation of a local project referred to as a 'Blue Wave Project'. Each of these projects had its own focus but with a common aim between them; making waterborne transport more attractive to shippers and hauliers and reducing the traffic nuisance caused by bridge openings. Alongside the realisation of these local projects the BGV platform was established.

The goal of BGV is to make IWW transport more attractive to shippers and hauliers and to reduce the traffic nuisance caused by bridge openings. This is achieved through the implementation of various local ITS-based projects and the establishment of an open data platform. Since the establishment in 2012, BGV has expanded to include eight governmental bodies as well as the National Data Warehouse. BGV now operates as a knowledge platform for all participants and a data platform for the data service.

In the BGV platform, the participants regularly meet, discuss progress on the local projects and exchange experiences and ideas. In addition to the meetings, BGV also included the establishment of an open data services. The data services includes data on the availability of berths and the current and estimated operation times of bridges. The data service is available online through https://blauwegolfverbindend.nl or accessible through an API.

The four BGV participants were the initiators of the BGV data-service and signatories of the BGV cooperation agreement in 2012. The founders of BGV are shown in *Table 20*. The cooperation agreement has been modified and extended in 2015 and later on in 2018.

Since 2012, Rijkswaterstaat has taken up the project management and promoted a nation-wide implementation of BGV. Other governmental authorities joined the BGV platform by participated in the knowledge platform. This also involved participating in the data service and recording the necessary data and importing the data into the BGV system. The Port of Amsterdam, the province of Groningen and the National Data Warehouse Authority joined as well as signed the cooperation agreement. Other regional authorities joined but without signing the cooperation agreement, involving Zwolle-Kampen region, Twente region, Stedendriehoek (involving the cities of Zuthpen, Deventer and Apeldoorn), Groningen-Assen region, province of Zeeland and the province of Friesland. Although having not signed the BGV cooperation agreement, these organisations are also involved with the development of BGV.

Table 20 Current lead partners of BGV and their role within the data service

Organisation	Type	Employees	Role
Rijkswaterstaat	National body	8.700	Data supplier, data facilitator, project manager
Province of North-Holland	Province	1.400	Data supplier
Province of South-Holland	Province	1.600	Data supplier
Port of Rotterdam	Port authority	1.200	Data supplier

7.1.2 Sector/Mode

The information disseminated through the Blauwe Golf Verbindend (BGV) platform serves two transport modes, i.e. road transport and inland waterway transport, both for freight and leisure transport.

7.1.3 Case study motto

Blauwe Golf Verbindend presents real-time data on the current and estimated time of bridge operations and the availability of mooring places – information that helps skippers and road users to better plan their journey and arrival, with less hassle and annoyance and without unnecessary fuel consumption and exhaust emissions.

7.1.4 Executive summary

BGV is an open data service that publishes three types of data, data on the current status of bridges, data on the estimated opening time of bridges and data on the availability of mooring places. As of 2019, the data service contains the real-time information of over 180 bridges, of which 100 bridges with estimated times, and over 1,100 mooring places in the Netherlands. These numbers continue to expand as more bridges and mooring places are added to the system and more organisations are cooperating. In 2019, the data service includes data on a number of locks as well.

The platform was launched in 2012 and established through the combined effort of three governmental bodies. The Dutch infrastructure manager *Rijkswaterstaat* is now national project manager, hosts the BGV data service and supplies data to the system. Data was initially supplied by the three participating regional authorities, the Dutch provinces of South-Holland and North-Holland and the port authorities of Rotterdam and Amsterdam. Since 2012, more regional authorities are providing data to BGV. The National Data Warehouse (NDW) offers support to the participants in data collection and disseminates the data through its own web service as national access point of traffic management data.

The data disseminated through the BGV platform serves different purposes. Information on mooring place availability is used by cargo skippers to search for a place to anchor their ship.

Information on the status of bridges can be used by skippers (professional and recreational users) and road users for route planning. Overall, the aim of the participants is to improve services to waterway- and road users through the use of innovative technology, eventually increasing the predictability, reliability and efficiency of waterway and road transport.

Figure 23 Overview of the number features bridges or mooring places in the Netherlands (Screenshots from the BGV web portal, available at www.blauwegolfverbindend.nl)

Figure 24 Data on a selected mooring place (in figure "Afrikahaven Wachtsteiger 1" in the port of Amsterdam), and other available data on the berth place or a bridge

Figure 25 Screenshot showing data on the operation times of a selected bridge ("Beatrixbrug, Zaandam'), near Amsterdam

The establishment of the BGV data service is strongly related to four other regional projects, called the 'Blue Wave'-projects. The four projects, described in *Table 21* are carried out by an individual participant within its own region. These projects started after receiving 50% cofounding by the national government through the national grant program '*Beter Benutten'* (in English: 'Making Better use of'), launched in 2011. While each project has its own focus, there was a common aim between them; making waterborne transport more attractive to shippers and hauliers and reducing the traffic nuisance caused by bridge openings. Cooperation between the participants and development of BGV was part of the grant application. To better facilitate their common goal, the participants decided to combine the data that is generated through these projects within a single platform.

Table 21 The four Blue Wave projects that led to the establishment of BGV and the implementing authority

#	Project	Participant
1	Generating real-time travel information for water and road transport for better route planning	Province of South-Holland
2	Coordinating bridge openings of different bridges for better traffic flows	Province of North-Holland
3	Developing a Bridge Management System (BMS) for supporting bridge operators in determining the best time to operate a bridge	Province of North-Holland
4	Developing the Inland Waterway Berth Information System for disseminating information on the availability of berth places	Port of Rotterdam

In the following years, more regional authorities joined BGV. In the meantime, the national government decided to extend the '*Beter Benutten*' grant program in 2014, triggering other regional authorities to initiate a local Blue Wave project.

The real-time data is made available to third parties through a web services. Rijkswaterstaat hosts the main BGV database, the central hub to which the participants send their data. Data relevant for road users are transferred to a web services that is hosted by NDW. In creating this web service, the participants invite third parties to develop water- or road-based route planning software that incorporates the data collected through the BGV platform.

Currently, as an example, BGV data is used in the water-based route planning application called *RiverGuide*[106]. For road traffic, the data is used by *Flitsmeister*[107], a route planning software, and the Intelligent Transport Services (ITS).

7.2 TECHNICAL PERSPECTIVE

7.2.1 Data sources/Uses

The BGV data service connects data from multiple sources. BGV involves three types of data; data on berths, the current status of bridges and the estimated operation time of bridges. Each of the data types is described below.

7.2.1.1 Availability of berths

Data on the availability of berths is provided through an information system called the *Binnenvaart Ligplaats Informatie Systeem* (BLIS, in English: Inland Waterway Berth Information System). BLIS is the result of one of the four 'Blue Wave-projects' that was carried out by the Port of Rotterdam. BLIS makes use of the Automatic Identification System (AIS), a transponder technology with which each inland cargo vessel is equipped. The database of information on berth places continues to expand as more and more organisations connect their berths to the system. In 2016, the Port of Amsterdam registered its berths in the system and since 2018, Rijkswaterstaat, the owner of most public berths on the main fairways in the Netherlands, connected their locations to the system as well. Other organisations are following. Adding more mooring places to BLIS requires the cooperation of the owners of the mooring place, which in many cases are municipalities or private parties.

7.2.1.2 Bridge status

Data on the status of bridges are initially collected by Rijkswaterstaat, the province of South Holland and the province of North Holland. To record 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.

7.2.1.3 Estimated bridge operation time.

Data on planned bridge openings is currently only collected by the province of North-Holland. The province has developed a system, the Bridge Management System (BMS), which provide

[106] RiverGuide application, accessible on http://riverguide.eu
[107] Flitsmeister application, accessible on https://www.flitsmeister.nl/

bridge operators with real-time information to determine the best moment to operate a bridge. The system, currently in its third version, combines data from multiple sources, including road traffic data, water traffic data, emergency services data and public transport data – all real-time data. Based on this data, the bridge operator determines what the best time is to open the bride for ships to pass. This data is then sent to the BGV database.

Bridge Management System (BMS)

One of the components of the BGV data ecosystem is the BMS. This system is developed by the province of North-Holland for its local Blue Wave-project. BMS combines many data sources and data flows, which are shown and collected within a user interface. The data sources involve: weather information; road and waterway maps; AIS data; GPS information (from the RiverGuide application); passageway registrations; data on road and waterway disruptions; live data on current road traffic (from NDW), public transport (from CROW) and emergency services (from p2000) and data from the local traffic management system. Based on the data shown in the user interface, the bridge operator plans the best time to open a bridge. The planned opening time is then send to form BMS to BGV. The user interface is shown in *Figure 26*.

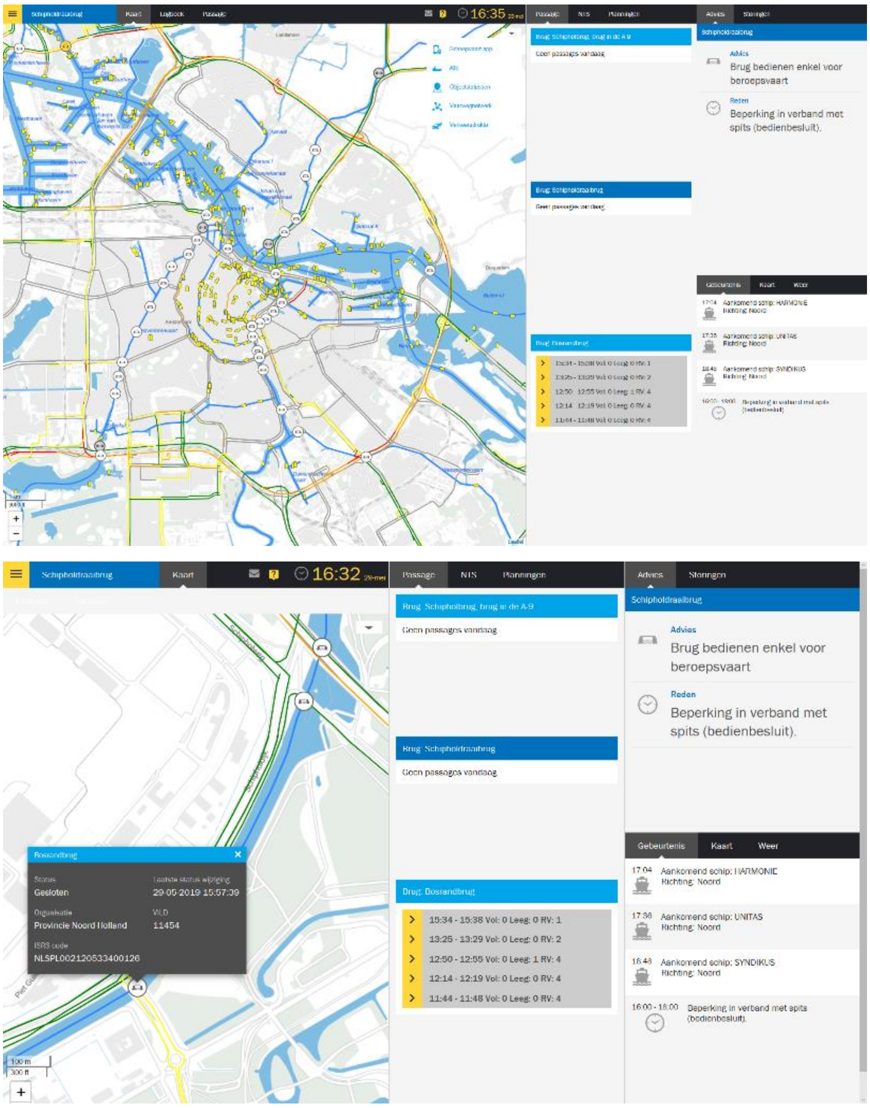

Figure 26 Screenshots from the user interface of BMS version 2.0 that is used by bridge operators. BMS is currently in its third version. Source: Province of North-Holland

7.2.2 Data flows

The BGV data flow consists of five main interfaces, which are explained in more detail below.

- *Interface between bridges and BGV*. Information on the status of bridges are mainly either recorded via industrial control systems and directly sent to the BGV system or through LoRa sensors, which signal is processed on a local server and then passed on to the BGV system. The province of North-Holland supplies data that is collected through its BMS.
- *Interface between mooring places and BGV*. Through BLIS, data on the availability of mooring places are sent to BGV. BLIS contains information on the geographical location and size of a mooring place. The system notices when a vessel occupies a mooring place based on the AIS signal that each cargo vessel broadcasts. If the location shown by the AIS signal is aligned with the location of the mooring place, the system records the berth to be occupied. The length and width of the ship are derived from the AIS signal. The occupancy rate is then calculated with the data stored in the BLIS database. An issue is when the AIS signal stops transmitting, for example when the skippers leaves his ship and turns of the power. In that case BLIS keeps the mooring place occupied until the AIS signal will be broadcasted again and a new calculation is possible.
- *Interface between BGV database and NDW database*. Data from the BGV system is transferred to the database of NDW using the DATEX2-format. DATEX2 is a European data standard for exchanging traffic related data between organisations.
- *BGV web service*. Third parties can request access to the real-time BGV data through a web service that is hosted by Rijkswaterstaat. The data flow can be set up through an Application Programming Interface (API) and is accessible to anyone. The data flow consists of real-time data on mooring places and the bridge status. Since 2019, the service contains data on a handful of locks as well.
- *NDW web service*. Similarly, to the BGV web service, third parties can request BGV data through NDW. This web service only contains data on the status of bridges, but may be favoured over the BGV web service as NDW offers other real-time traffic data, such as data on road constructions, congestions or accidents.

7.2.3 Main technological challenges and issues

BGV involved new technology, meaning that setting up the system was a challenge. The BGV data service had to be constructed from scratch and all participants needed the technology and software to send data to the BGV data service. This also meant the development of data standards and ensuring continuity of the system. An enabler was that the participants could build on the data technology skills that were already available to them within their organisation. New hardware was required to translate the physical closing and opening of bridges into digital data. The NDW, the data facilitator, has experience with data collection and facilitated the participants in purchasing and installing the hardware required for registering changes in bridge status. For the participants, supplying data to BGV meant connecting local systems to the BGV server.

Using new technology also meant that the persons using the technology needed to learn and adapt to the system. This was particularly challenging for the bridge operators, who needed to keep track on many data sources and indicate the estimated times of bridge openings. This also applied to the installation and maintenance of software installed on bridges.

No data analytics such as machine learning algorithms were part of the project. Hence, knowledge on the technologies related to data analytics was not required for BGV. Nevertheless, all participants expressed interest in developing BGV from disseminating real-time data towards publishing predictive analyses.

Another recurring theme in the big data literature was the difficulty of finding skilled data scientists (Miller, 2014). According to the technical expert of Rijkswaterstaat, this was not an issue for the development of the BGV platform. While it was acknowledged that governmental organisations such as Rijkswaterstaat cannot compete with the salaries that are being paid by commercial parties, there are various reasons why this does not apply to them. The interviewee mentions that (i) the societal relevance of the work, (ii) the variety in the work and (iii) the size and uniqueness of the data available, are attractive conditions for data workers to work for public organisations. Large governmental organisations such as Rijkswaterstaat have the data and the projects that compensate for the lower salary.

While the technology was a major component of BGV, the participants consider the main challenge of BGV to be gathering sufficient the political will behind the project.

7.2.4 Big data dimensions and assessment

The literature distinguishes four dimensions of big data: volume, velocity, variety, veracity. Each of these four components is elaborated in more detail in *Table 22*. For BGV, we conclude that whereas the volume and variety of the data is relatively small, the velocity of processing real-time data poses more work, with the veracity of the data being the most challenging component.

Table 22 The four V's related to BGV

Component	Relevance to BGV
Volume	With around 180 bridges (and 12 data points per bridge) and over 1100 mooring places connected to BGV, the volume of the data processed is relatively small. BGV looks at real-time data, meaning that only a short duration slightly before current time is made available through the data service.
Velocity	While BGV primarily involves dynamic data, the amount of objects that are broadcasting real-time data is limited. Approximately 1300 objects are currently connected to BGV, broadcasting real-time information into BGV. The permitted latency between the event (e.g. changing status of a bridge) and the registration of the change in the NDW database is agreed between the participants to be 1 second.
Variety	The variety of data involved with BGV is fairly limited. BGV involves three types of data: data on berths, bridge status and planned bridge opening.
Veracity	From the four V's, data veracity is the most challenging component. Reliability issues for berth place data occur when changes in length, width or cone status of a berth is not

	recorded in the system, or when skippers do not fill in the correct width and length of the ship,. However, this was not considered a big issue by the interviewed skipper nor did the port of Rotterdam received feedback of issues regarding the reliability of the data. An evaluation of the Blue Wave-project in North-Holland showed that skippers considered the reliability of the estimated bridge opening to be insufficient. This aligns with the difficulties involved with estimating the right time. For that reason, the participants agreed on the reliability norm for this data to be set at 80%. Nowdays the reliability of the estimated opening time is improved. Fewer issues are observed with the real-time status of bridges, for which the reliability norm is set a 95%. But there is still room for improvement considering this data is used for route planning. A situation where 5% of the road trips are being detoured when there is in fact no closed bridge can add up to be substantial. To increase data quality, various checks are built in by Rijkswaterstaat and NDW to control the accuracy of the data received by the participant, such as: Only one 'open' signal can be sent per bridge For every 'open' signal, a 'close' signal needs to be broadcasted Bridges are checked if they are open longer than 15 minutes. Keep alive checks are check if the bridge signal is available Improving the data quality is a continuous challenge for BGV.

7.3 BUSINESS MODEL

The BGV project consists of two levels of business models.

- The first level involves the business model of each of the participants separately. Each of them has their own motivation and vision for participating in BGV. Participation is based on an individual business model.
- The second level is the business model that is shared between the participants. Understanding the first level is perquisite for fully understanding of the second level.

Individual business models

The first-level business models are summarized in *Table 23*. The provinces are involved for traffic management purposes. Through disseminating the opening times of bridges in their province, their aim is to limit the hindrance open bridges is causing to road- and waterway traffic. Evaluation shows a reduction of 2 million euro social costs caused by less waiting time by bridges in the province of North Holland. For the port authorities, their aim is to disseminate information on the availability of berth places in their port. This helps the skippers in their search for a place to anchor their ship. The involvement of Rijkswaterstaat and NDW is more facilitative. As the national infrastructure manager, Rijkswaterstaat is supporting initiatives that make the transport more efficient on a national level. The task for NDW is to facilitate the process of data collection and dissemination.

Table 23 Business models of each of the BGV participants

Aim	Description	Main actors
Traffic management	limiting the disturbance caused by open and closed bridges to road and waterway traffic	Provinces of South and North-Holland, RWS
Service provision to skippers	Provide information on the availability of mooring places to skippers	Port authorities of Amsterdam and Rotterdam, RWS
Transport improvement	Improving Dutch infrastructure	Rijkswaterstaat
Data dissemination	Improving access to road traffic data	NDW

BGV business model

The business model canvas as shown in the figure below is used as the basis for describing the business model behind BGV.

<table>
<tr><th>Key partners</th><th>Key activities</th><th>Value Proposition</th><th>Customer relationships</th><th>Customer segments</th></tr>
<tr><td rowspan="3">Owners of berths and bridges; third parties implementing BGV data for customers segments to use</td><td>Collect reliable data; set up reliable web service; expand BGV database</td><td>Societal value by delivering data that helps skippers and road users better plan their jour</td><td>Primarily through RiverGuide and service providers</td><td rowspan="3">Waterway- and road users (freight & passenger, leisure and professional)</td></tr>
<tr><td>Key resources</td><td>Internalization of costs and benefits</td><td>Channels</td></tr>
<tr><td>Collaboration with key partners</td><td rowspan="3">No relevant external costs are associated with BGV</td><td>Through third party route planning software</td></tr>
<tr><td colspan="2">Cost structure</td><td colspan="2">Revenue streams</td></tr>
<tr><td colspan="2">Maintenance and marketing</td><td colspan="2">Users are not directly charged (monetary or through information), but they contribute indirectly through taxes</td></tr>
</table>

Figure 27 Business model canvas of BGV

Each of the ten components is elaborated below:

- *Value proposition*. BGV unlocks real-time data on the status of bridges and berths available in ports - information that helps skippers and road users better plan their journey and arrival, with less hassle and annoyance and without unnecessary fuel consumption and exhaust emissions.
- *Customer segment*. Targeted customers are any road- and waterway users, both for passenger transport and freight transport, and leisure and professional users.

- *Customer relations*. In the development stage of BGV, customers were consulted through the smartphone application RiverGuide. Apart from marketing the RiverGuide app, currently little interaction with customers takes place
- *Channels*. The customers are reached through third parties offering route planning software. BGV contains a free web service through which third parties can access the BGV data. Flitsmeister, Waze and RiverGuide are applications that are making use of BGV data. In addition customers are reached through websites that publish the BGV data, such as its own website, www.blauwegolfverbindend.nl, or third party websites such as www.brug-open.nl, www.isdebrugopen.nl or https://brugopen.nl. However, the main use of BGV is its integration in route planning software. BGV does have contact on a regular base with the service providers.
- *Key activities*. Key activities to deliver the value proposition are collecting data on bridges and mooring places; ensuring the reliability of the generated data; transferring the data from the participants to the BGV web services of Rijkswaterstaat and the NDW web service; having third parties use the BGV data and deliver it to users; and persuading more regional authorities to link op to BGV.
- *Key resources*. As the infrastructure for collecting and transferring data is established, this system does not require significant resources apart from maintenance and optimisation to run. The cooperation of local owners of bridges and mooring places is key in maintaining and expanding the BGV data.
- *Key partners*. Key partners are those that deliver bridge and mooring place data (i.e. local owners, primarily municipalities and provinces). BGV benefits from the networking effect; the more information is available, the more value the system delivers. Hence, the system benefits from more organisations participating.
- *Cost structure*. The bulk of the costs were made in setting up BGV. Half of these costs were subsidized. Now, each year, only a small portion of the costs is needed to operate and maintain the system.
- *Revenue streams*. No direct compensation is paid for the BGV data, nor do the users provide their data as consideration for accessing the service – as is often the case in big data based business models. The revenue is generated through more efficient and reliable traffic systems, in monetary costs and emissions. As public institutions are behind BGV, the project is more about societal value than short term economic value.
- *Internalisation of costs and benefits*. The data published through BGV promises to help with more economic sailing and reducing congestion for bridges. This has positive effects on greenhouse gas emissions. Moreover, making inland waterway traffic more attractive contributes to modal shift, which is automatically means reduced greenhouse gasses. To what extent this is the case has still to be researched. A pilot evaluation study on traffic management based on regulating the opening times of bridges in the province of North-Holland suggests that cars spent approximately 10% less time in traffic (Arcadis, 2018).

7.3.1 Business processes

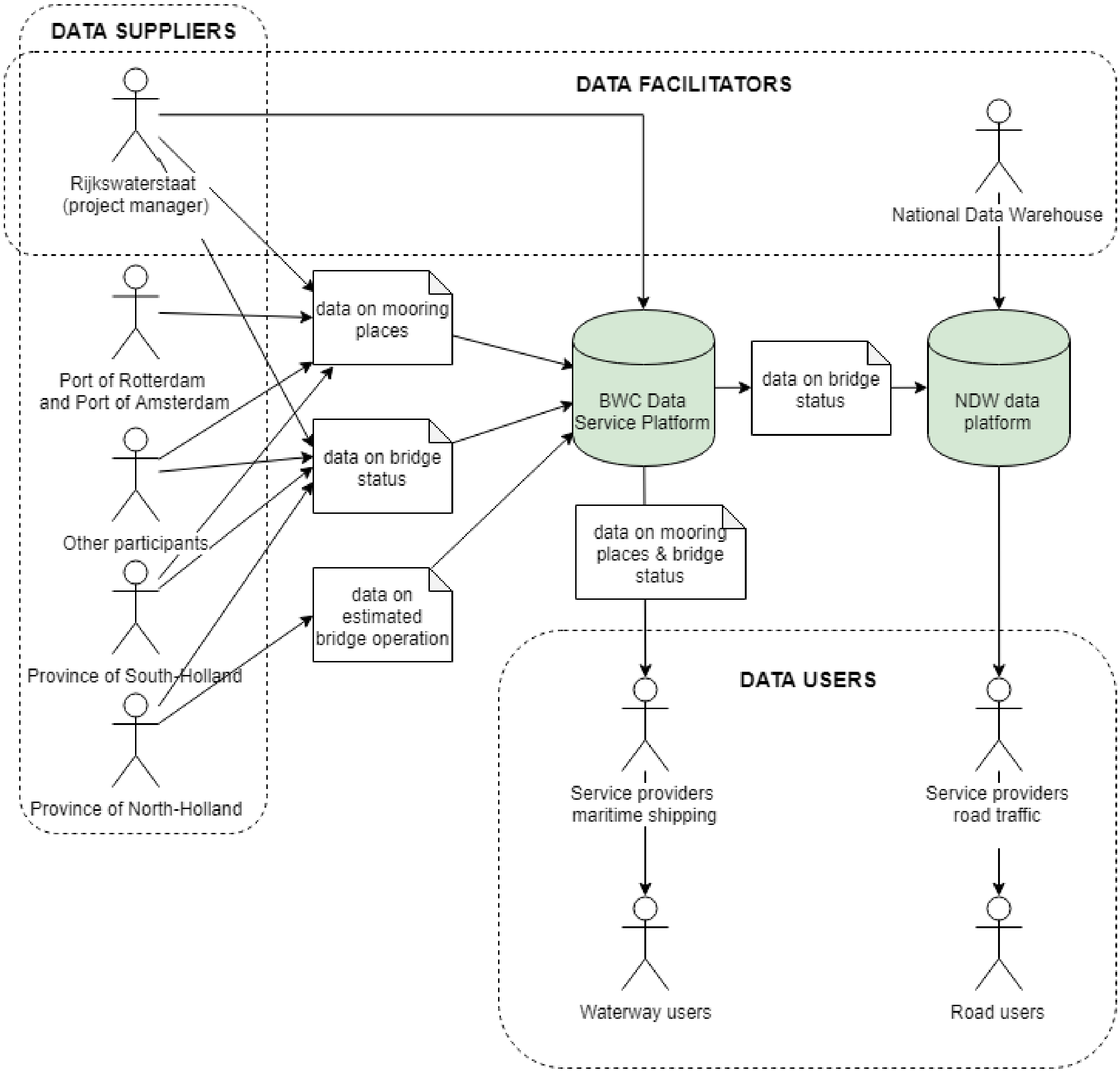

Figure 28 The BGV business process, showing the organisations involved, their roles and the data flows between the stakeholders

The business process is described in the figure above. The business process can be summarized in the following four steps

1. Data on mooring places and the status of bridges is collected by the participating organisations. The data is processed and stored on local serves.
2. Participants' data is transferred the BGV database that is hosted by Rijkswaterstaat
3. The data on the bridge status is sent to NDW
4. The information is made accessible through the BGV and NDW web service

7.3.2 Relation to big data initiatives

Each of the data suppliers developed a system to collect, process and store the data. The participants link their system to the central BGV system and feed data to the BGV database.

7.3.3 Illustrative user stories

User story one: skippers searching for an available mooring place

The skipper arrives at the port at the agreed time to deliver cargo. After the cargo has been unloaded from the ship, the skipper checks the RiverGuide application to choose the nearest free berth with good facilities visible in RiverGuide. Being able to check berth place availability online means that the skipper no longer has to sail into different ports in search of an available berth, which may also do not have facilities to come ashore.

User story two: BGV data used for route planning for road users.

A driver enters the destination in the route navigation system. During the journey, the driver receives a new route proposal because a bridge opens on the current route. After having taken the alternative route, the driver arrives at his destination 2 minutes later than the original route, however, is spared being stuck in traffic for 10 minutes waiting in front of an open bridge.

User story three: BGV data used for bridge operations and sailing.

A skipper enters his/her destination into the RiverGuide App. When starting to sail, the app informs the skipper that the bridge operator has noticed the skipper. Based on other waterway traffic and road traffic, the bridge operator determines an estimated opening time of a bridge. The opening times of multiple bridges on trajectory are coordinated by the bridge operators to create a so called 'Blue Wave', i.e. the bridges are synchronized in such a way to ensure a continuous flow of traffic along the whole stretch of the waterway. The estimated opening time is sent back to the skipper. The skipper slows down a little, saving fuel, and sails quietly, together with two other ships, in order to arrive at the bridge just in time as it opens for the skipper. Because the three ships stay together in a convoy, the bridge needs to open less often and road traffic stops less for open bridges.

7.4 ANALYSIS OF SOCIETAL ASPECTS

The societal aspects of BGV are identified on the basis of the aspects identified in the literature review conducted under work package two. Note that some issues overlap between the aspects. For example, privacy related issues are at the same time legal issues (due to the existence of legislation related to privacy and data protection) and social issues (as the need to have privacy-related legislation among others derives from the negative societal consequences related to the spread of potentially sensitive personal data).

7.4.1 Economic and political

The economic value from BGV is derived from more optimized route planning. Road users spend less time waiting for open bridges. Waterway users save time in searching for a place to berth and knowing the operation time of bridges enables them to sail more economically. The full extent to which this is true in practice has not been researched yet.

Data on the effects of BGV on traffic is limited by one evaluation conducted in the province of North-Holland (Arcadis, 2018). The study attempts to quantify the effects on the regulation of bridge operation times. The study is limited by the short timeframe over which traffic data is collected and does not account for the differences in opening times of bridges. Considering this, the study finds that the time vehicles are waiting for closed bridges is reduced by 12% during weekdays and 9% during weekends.

Another issue is the dependency of BGV on third parties for the data to reach its customers. BGV needs to be implemented by third parties into route optimisation software so that can be used by road- and waterway user. At the same time, the open data nature of BGV enables third parties to innovate with the data to their liking, giving the opportunity to markets to create their own products.

With waterborne transport being made more efficient, another benefit could be modal shift from road to water. However, the respondents think that the effect of BGV is too limited to cause a modal shift.

In addition to the direct economic value, BGV is the first step to create an inter-organisational database with data on inland shipping. The current collaboration and associated trust-building between the organisations enables future collaboration, for example by adding more data to BGV, inviting more participants, providing extra data services, perhaps even data analytics or more collaboration in traffic management. This is particularly promising since BGV concerns data-based technologies for which there are many potential applications.

7.4.2 Social and ethical

An analysis of the societal and ethical issues is done according to the six topics identified by (Debussche, César, Hong, et al., 2018): trust, surveillance, privacy, free will, personal data ownership, discrimination and environment.

Table 24 Overview of societal and ethical aspects and relevance to BGV

Aspect	Description of aspect	Relevance to BGV
Trust	Increase of decrease in trust in by or between the data users, data facilitators and data suppliers.	The cooperation between the different governmental organisations runs smoothly, so that high amounts of trust is built between the participants. This eases the process of expanding the services of BGV, which the participants aim to do. The reliability of the data on planned opening of bridges should be increased, so that more skippers and road users are willing to use this service.

Surveillance	The ability to collect large amounts of data enables the data user to better understand processes in society.	The centralisation of data by BGV data was not done before and offers new insights in traffic management. Since the data is open, others have access to this data as well.
Privacy	The involvement of privacy sensitive information	While AIS-data is classified as personal data, the data is only available for traffic management purposes and thus in possession of the participants anyways. The AIS data is not available as open data Moreover, agreements are made between the shipping sector and the government about the situations when the government can use this type of data. For example, it is agreed that AIS data cannot be used for law enforcement. However, the shippers do note that they are not willing to give up AIS-data for particular uses when there is nothing in return. This is currently not the case, but may pose challenges for further development. It took 3 years for Rijkswaterstaat to have the usage of AIS-data internally approved. Hence privacy is mostly an issue for the BGV participants as users of data being required to comply with privacy and data protection-related legislation, rather than being an issue of data suppliers seeking to assert their privacy-related rights.
Free will	Big data analysis and associated predictive analysis that threatens human autonomy	While the traffic managers aim to control the flow of ships, the skippers encourage such management, as it enables them to save fuel and be more punctual. BGV data is now used for predictive analysis. However, veracity is the most challenging component.
Personal data ownership	Confusion of who owns the data or loss of data ownership by data suppliers	The data produced by BGV is open, making data ownership of BGV data not a topic of discussion. Due to agreements between the shipping sector and government about the usage of AIS-data, the ownership of this data is not considered an issue.
Discrimination	Discrimination of population groups based on generalisations derived from big data	No issues related with discrimination were found.

In summary, only few social and ethical issues were found. First, the data mainly concerns vessels instead of people and other non-personal types of data, thereby eliminating many issues such as discrimination and threats to free will. Second, the data of BGV is mostly used for informative purposes. Therefore, with no generalisations or analysis being made, there is no threat of a sense of loss of autonomy. Third, because the organisations behind BGV involve governmental organisations, societal and ethical issues are usually filtered out or aggregated before the data service is published. This is evident through the usage of one of the used data sources for BGV, AIS-data, which is privacy sensitive. Through negotiations with the shipping sector about the usage of the data, in combination with the public organisation doing significant effort to comply with current privacy legislation, it is made sure that the data is clear of social and ethical issues before being released.

7.4.3 Legal

It is important to note that big data policies are laid down in the cooperation agreement signed by the participants. The document describes the organisation of BGV, the tasks of the participants and how a reliable service is assured. The current cooperation agreements entered into force in 2018 and are the follow-up and expansion of the previous cooperation agreement signed in 2012 and extended in 2015.

To analyse the legal issues, the framework outlining the legal issues relevant to the production of, access to, linking of and re-use of big data in the transport sector (Debussche, César, & De Moortel, 2018) is used. They identify, among others, the following legal dimensions of big data:

Table 25 Overview of legal aspects and relevance to BGV

Aspect	Description of aspect	Relevance to BGV
Privacy and data protection	Consideration and implementation of the requirements of the GDPR and other types of privacy and data protection-related legislation	Participants underwent significant efforts to ensure compliance with the various requirements introduced by privacy an data protection legislation
(Cyber-)Security	Awareness and implementation of certain general or specific (cyber-)security requirements	Participants could draw on IT knowledge and systems within their respective organisations. The fact that most data used and published in the context of BGV is open data also has consequences and eliminates a certain number of security threats, particularly those related to a loss of confidentiality of the data.
Anonymization and pseudo-nymisation	The use of any type of anonymisation or pseudonymisation technique, i.e. techniques	The participants have undertaken significant efforts to anonymise, to the largest extent possible, AIS-data before making it available.

	to ensure personal data can no longer be attributed to individuals without using additional information or to reduce the identifiability of individuals	
Supply of digital content and services	Uncertainties regarding supply of digital services for personal data	Since AIS-data is already freely available, it is not considered as a 'payment' for BGV data
Intellectual property	Protection by the organisation of its service/product through intellectual property rights and hindrance for the organisation of third parties claiming or otherwise asserting intellectual property rights	The data is available as open data. With public organisations involved, future claims of intellectual property are unlikely.
Open data	Awareness and use of open public sector information	BGV relies on open data. Furthermore the Dutch government is actively providing open data to the public. It has adopted an 'open unless' default approach and is focused on providing as many datasets as possible. Currently more than 7,000 datasets on different subjects are freely available on the internet among which traffic data. For example: Current traffic data gives an up-to-date picture of the situation on the road. Every minute, data from more than 24,000 measurement locations in the Netherlands are processed in a database and distributed to customers. It concerns data about intensities, travel times, point speeds and vehicle categories. All data produced in the context of BGV is in made available to the public for free and as such must also be considered open data. In general, the Netherlands has adopted several acts in relation to making data open, accessible and re-usable. With regard to re-use of (big) data, The Netherlands has transposed the Directive 2013/37/EU into the Act hergebruik van overheidsinformatie (*reuse of government information*). This law provides options for citizens and companies to submit requests for the provision of government

		information. Information and data from the government should be actively made available in a machine-readable format to be reused for other purposes, even commercially. The main goal of this law is to create economic added value.
Data sharing obligations	Awareness of and compliance with certain specific data sharing obligations, mandatory legal requirement to share data with other market actors	BGV relies on data sharing of all participants. The sharing of data and its quality are laid down in a cooperation agreement, however, not legally binding.
Data ownership	Third parties claiming data ownership	The data is available as open data. With public organisations involved, future claims of data ownership are unlikely.
Data sharing agreements	Use of data from non-public organisations and the organisation of the data sharing	Agreements are made between the shipping sector and the government about the situations in which the public bodies can use this type of data. For example, it is agreed that AIS data cannot be used for law enforcement.
Liability	Consideration of legal liability framework and any particular issues in this respect	BGV takes no responsibility for use of the data it makes available. Any use of BGV data is the user's own responsibility. Moreover, considering the nature of the data made available, the consequences of wrongly displayed information are not likely to be significant, making liability claims unlikely.
Privacy and data protection	Consideration and implementation of the requirements of the GDPR and other types of privacy and data protection-related legislation	Participants underwent significant efforts to ensure compliance with the various requirements introduced by privacy and data protection legislation

In summary, few legal issues are encountered as part of BGV, the fact that BGV uses open data and in turn makes the data available freely as the most significant legal aspect. Given that the data used and produced as part of the BGV project is for the most part not considered personal data or otherwise of a sensitive nature, the legal issues in this case study are limited. In addition, the participants had done significant effort to comply with legal requirements regarding privacy, including through the use of anonymization techniques.

7.4.4 Environmental

Environmental effects are described as 'the difference between the expected and observed environmental impacts from new technologies aiming at efficiency improvements' (Andersen & Benkic, 2018). Additional waterway traffic is often considered to be positive, as it indicates a shift from the more dangerous and environmentally harmful road traffic, while waterway congestion is not an issue. A bigger concern for BGV is the extra road traffic that a more optimized route planning system generates. Reduced travel times and increased travel time reliability generates new road traffic that is attracted by the shorter travel time. However, no study has been conducted yet on the reduced travel times resulting from optimized route planning with BGV data. Moreover, these indirect effects are usually only visible in the long term.

7.5 ASPECTS RELATED TO OPPORTUNITIES, BARRIERS AND LIMITATIONS

Summarized in the table below are the opportunities and challenges of BGV.

Table 26 Opportunities, enablers, barriers and limitations of the BGV

Opportunities and enablers	Barriers and limitations
Profit for road and waterway users. BGV helps route optimisation by delivering real-time information of bridges. This limits the time road vehicles are waiting for bridges. Data on the availability of berth places limits the skipper search time for a place to moor and saves fuel.	*Organisation*. The reliance on many organisations cooperating has delayed the realisation of BGV, in particular since participation was not always high priority for all organisations. For example, bridge ownership in North Holland is spread over more than 15 organisations, and disclosing the real-time data of these bridges requires participation and agreements with of each of them.
Transferability. With the main data infrastructure being developed, it is relatively easy for other organisations to connect their bridges and mooring places to the system. In doing so, the amount of data published by through BGV can be expanded easily.	*Different ambition*. The participants have different paces in further developing BGV. This is the cause of the different business models from which the participants are cooperating. While some want to focus on fine-tuning the current system (e.g. offer higher reliability), others would like to develop the current system further (e.g. include predictive analysis).
Scalability. BGV lays the foundation of an inter-organisational collaboration platform. The participants have gained support for the project within their organisations and have built trust and procedures for further cooperation. Together with ambition to further develop the	*Privacy*. The need to comply with privacy legislation has slowed down the development pace of BGV. It is moreover perceived as potentially obstructing further development of the platform. For the future, the participants expressed the desire to make more use of AIS-

platform, this is fertile soil for more cooperation. With the foundation being there, the entry costs for new participants are low. Experience, know-how and best practices are available for other governmental authorities to profit from.	data. However, this is associated with privacy-related issues. The trade-off between the desire for privacy and the added value of using personal data could be a major obstacle for further development of the platform.
Open data. The BGV data is open to anyone, enabling public and private parties to use the data for innovations.	*Long-term benefits*. The long-term benefits of road route optimisations are unclear due to the induced demand caused by shorter travel times.
Subsidy. It is highly unlikely that the project could take off without the initial subsidy, therefore making subsidy a crucial enabler for BGV	*Practical use*. BGV is a data service and not directly an end-product. The data is used by the provinces and Rijkswaterstaat for traffic management. But part of the data finds its utility when being implemented by third parties, such as Flitsmeister.
Data technology. No major challenges were encountered in setting up the data infrastructure, which offers perspective for similar big data initiatives. Whereas the volume and variety is relatively small, the velocity poses more work, with veracity being the most challenging component for BGV to be valuable. It helped that the participants were large organisations with experience and resources to handle such data technologies.	*BGV service has room for improvement.* BGV offers only a small portion of all the information that skippers need while navigating. Therefore, an argument raised by skippers for not using the data is that it 'requires another program', in addition to the existing programs that they are using. Moreover, skippers would like to see the quay length occupancy in the mooring place data. However, privacy plays an issue here to include this kind of data. Disseminating this data makes it easier to identify a skipper.

7.5.1 Concluding remarks

BGV is best considered to be an information source: an information source for data on the availability of berths and the current status of bridges. BGV is not a stand-alone big data application. The utility of BGV is achieved through third parties that decide to make use of and innovate with the BGV data. Hence BGV is equipped with an API that enables easy extraction of BGV data by third parties.

This facilitative function fits with the reasons that led to the development of BGV and the purpose of BGV. BGV is a combination of various local projects involved with data on smart inland shipping. The participants felt that combining this data in a central place could be a promising foundation for a (nationally) centralized data collection system for data related to inland waterways. No clear goals were defined at the beginning of the project nor did each of the participants know where this would be heading. After positive responses from data users and other regional authorities joining BGV with their data, the cooperation agreement signed by the participants in 2012 was extended and broadened in 2018 with the same premise; to improve services to waterway- and road users through the use of innovative technology.

The future direction of BGV is multiple. It will at least focus on increasing the reliability of the current service and continuing to expand the data coverage. In the long term there are many directions in which BGV can go; adding additional data categories, expanding the system internationally or expanding the data service with predictive analysis (for example, displaying the chance that a mooring place may be occupied, based on historical data and current waterway traffic).

At the same time, regional Blue Wave projects are being developed throughout the Netherlands. The Twente region is currently developing a whole new data system for inland waterway users, including data such as the availability of quayside electricity and even promising a market place for carriers to combine cargo. The technologies used within these regional projects could be interesting to adopt on a national scale within BGV.

BGV is thus more than a data service, it is equally a cooperation platform for regional authorities, traffic managers and data users to create a data-based system aimed at providing a better service to skippers and road users.

Another important conclusion is related to big data technology. BGV highlights how the application of big data does not necessarily pose a technological issue, but more importantly could be an organisational issue.

7.6 SOURCES OF INFORMATION

7.6.1 Interviews

Several interviews were carried out facilitated by the BGV manager at the Rijkswaterstaat:

- Manager at Rijkswaterstaat
- Manager at Port of Rotterdam
- Manager at province of North Holland
- Technical expert at Rijkswaterstaat
- Inland waterway shipping interest organisation
- Skipper

7.6.2 References

Andersen, O., & Benkic, M. (2018). Report on trade-off from the use of big data in transport.

Arcadis. (2018). Evaluatie Blauwe Golf Vervolg.

Belcastro, L., Marozzo, F., Talia, D., & Trunfio, P. (2016). Using Scalable Data Mining for Predicting Flight Delays. ACM Transactions on Intelligent Systems and Technology, 8(1), 1–20. https://doi.org/10.1145/2888402

Debussche, J., César, J., & De Moortel, I. (2018). Report on legal issues.

Debussche, J., César, J., Hong, M., Nordeng, Z., & Waldenfels, R. (2018). Report on ethical and social issues.

Ghofrani, F., He, Q., Goverde, R. M. P., & Liu, X. (2018). Recent applications of big data analytics in railway transportation systems: A survey. Transportation Research Part C:

Emerging Technologies, 90(March), 226–246. https://doi.org/10.1016/j.trc.2018.03.010

Miller, S. (2014). Collaborative approaches needed to close the big data skills gap. Journal of Organisational Designs, 3(1), 26–30. https://doi.org/10.7146/jod.9823

Milne, D., & Watling, D. (2018). Big data and understanding change in the context of planning transport systems. Journal of Transport Geography, (November), 1–10. https://doi.org/10.1016/j.jtrangeo.2017.11.004

Tanis, J., Teoh, T., Aavik, A. P., Burgess, . Arnaud, César, J., & Russotto, R. (2018). Deliverable 1.2: Big data policies.

Teoh, T., Hindriks, I., Van Berne, B., & Op ’t Ende, P. (2018). Report on economic and political issues.

TURBLOG. (2011). Business concepts and models for urban logistics.

Zaman, I., Pazouki, K., Norman, R., Younessi, S., & Coleman, S. (2017). Challenges and opportunities of big data analytics for upcoming regulations and future transformation of the shipping industry. Procedia Engineering, 194, 537–544. https://doi.org/10.1016/j.proeng.2017.08.182

Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., & Chen, C. (2011). Data-Driven Intelligent Transportation Systems: A Survey. IEEE Transactions on Intelligent Transportation Systems, 12(4), 1624–1639. https://doi.org/10.1109/TITS.2011.2158001

8 Case study 6: "Optimised transport & improved customer service"

Using privately owned vehicles for transportation has, for long, been deemed as a convenience by commuters. However, increasing congestion on roads and emergence of shared mobility services have allowed them to explore alternative options. Still, a marked change in the behaviour of commuters has not been seen. In Belgium, for example, cars remain the preferred mode of transportation. This may be attributed to the fact that available alternatives (i.e. going on foot, by bike or by public transport) is slower by as much as 67%.[108]

In this context, the present case study on "Optimised transport & improved customer service" explores the role of big data-based applications that can help optimize mobility and improve consumer experience. Such applications can support both transport service providers as well as end-users by suggesting the most efficient routes based on real-time data feeds on traffic and public transport. They can, eventually, also prove useful to traffic controllers, helping them relieve road congestion. The case study is based on interviews with Nextmoov, a Brussels-based smart mobility digital agency.

8.1 CASE STUDY OVERVIEW

8.1.1 Organisation/Stakeholders

Nextmoov is a Brussels-based **smart mobility digital agency**. It started in 2012 as a **start-up** with the initial aim to build a mobile companion for public transport, similar to Citymapper[109] and Moovit[110], which was lacking in the Belgian market. The product they built is called **NextRide** and it provides real-time public transport schedules for both Brussels and Wallonia. It targets users of public transport but can also be used by public transport service providers, who can receive real-time data collected by the NextRide community. NextRide was built during a period of two years with the help of several 'incubators' and 'accelerators'. NextRide has had a lot of success; at one point one in three public transport users in Wallonia used it as their main application to get timetables and routing suggestions. It has also been downloaded almost 1 million times; a rather impressive number given the size of the market.

The NextRide application used to be handled by a team of 4 people working on it in their spare time and addressing up to 4 million user requests per month. This experience provided an opportunity for Nextmoov to learn how to use its limited resources in the most efficient way possible and this ability is something that it brings to its clients, also in its current projects.

As a start-up, Nextmoov stayed a 4-people company up until 2017, at which point it started to expand. Today, it includes approximately **30 people** in its team and is implementing multiple projects. In 2017, Nextmoov also made a strategic choice between trying to grow and attract money from investors (based on the example of City Mapper and other similar applications) OR

[108] Deloitte (2019). A new deal for mobility in Belgium quoting De Tijd (2019). The Congestion Report.

[109] Citymapper is a public transit app and mapping service which started in 2011. It integrates data for all urban modes of transport with an emphasis on public transport. It is headquartered in London.

[110] Moovit is a Mobility as a Service (MaaS) provider and free public transit app. The company was founded in 2012.

focusing on the Belgian market and **helping other actors in Belgium to improve their digital assets** (the way it did with NextRide). As a Belgium-based company, Nextmoov decided for the latter, keeping in mind **its main goal – to improve mobility in Belgium**.

Nextmoov addresses a large community in the field of mobility. Its products target end users of mobility solutions as well as integrators of solutions, media companies and mobility service providers. Operators of mobility services (such as shared bike, scooters, public transporters, shared vehicles etc.) are not only potential providers of data, but also users of Nextmoov technology. The beneficiaries of Nextmoov are primarily located in Belgium, but the intention is to also expand internationally.

In order to provide more integrated services, in 2018, Nextmoov has created a joint venture called "Stoomlink" with a mobility consulting firm (Stratec) and the Belgian Road Safety Institute (Vias institute). The joint venture aims to provide comprehensive solutions for mobility issues by providing both consultancy as well as technical support.

In terms of Human Resources, Nextmoov works only with **freelancers** and not all of them work full-time. This allows the company to be flexible and maximize efficiency. The freelancers are based all around Belgium. A major section of human resources is composed of **Javascript developers**, but it also consists of **graphic designers and project managers**. This work structure contributes to the company's success as each team member knows what they are working on and why, and all the principles adopted by Nextmoov express its overarching and comprehensive approach.

The work schedule is not divided in terms of 'working hours' and there is no specific 'office space' except for the Head Quarters of Nextmoov. The company takes pride in advertising itself as having the smallest office in Brussels, which it feels is another strong feature of its distinctive structure. Its value lies in being efficient and result oriented. One of Nextmoov's principles is to **automate as much work as possible** and it aims to be as paperless as feasible. It uses "Slack"[111] for internal communication and only holds physical meetings in small teams, when absolutely necessary. Its working language is French due to the predominance of French-speaking associates, but it intends to move towards English with a view to future expansion. Its coding language, however, is English, which is still not the case for many digital agencies in Belgium.

Currently, the Chief Executive Officer (CEO) of Nextmoov makes all important decisions of the company. But the intention is to establish a more horizontal hierarchy in the future. Nextmoov is also in discussion with lawyers about the **possibility of becoming the first company with Artificial Intelligence (AI) as the CEO**. It already uses an AI assistant to schedule its meetings and it plans to further build on this technology to automate project-related reporting, invoicing, etc. It aims for transparency in its internal communication and the teams are always involved in all relevant email exchanges at all times.

[111] Slack is a cloud-based set of team collaboration tools and services.

8.1.2 Sector/Mode

Nextmoov operates in the **passenger transport** sector, across all urban transport modes in Belgium: **rail, tram, metro and bus**. Its operations also touch upon modes such as bicycles and scooters, which have become more popular in recent times.

8.1.3 Case study motto

The first mobility planner combining real-time datasets from all major public transport operators (PTOs) in Belgium.

8.1.4 Executive summary

This case study focuses on the role of big data-based applications that can help optimize mobility and improve consumer experience. Accordingly, it focuses on the most recent product developed by Nextmoov - the Smart Mobility Planner (SMP), as an example for the case study.

SMP is particularly interesting and relevant considering the manner in which it has been funded and the approach that Nextmoov has adopted for developing this product. Nextmoov has been working with the four main transport operators in Belgium to operationalize SMP, which could be **the first route planner covering Belgium, merging real-time data feeds of the four PTOs in the country**, including, **TEC** (bus operator covering Wallonia), **De Lijn** (bus operator covering the Flanders), **STIB** (bus, tram and metro operator covering the Brussels metropolitan area), and **SNCB** (Belgian national train operator).

The closed alpha version of SMP was launched on 4 March 2019 and the open beta version is being released in May 2019. The testing phase involves 4,000 test users and is expected to be concluded by summer 2019. The test users were recruited via online social media campaigns ran by PTOs. The next step will be to implement the solution in the back-end of the PTOs. Each PTO will keep their own (improved) application under their individual brand and the names of their respective mobility planners. As a result, the SMP will only exist as an individual application for the testing purposes.

8.2 TECHNICAL PERSPECTIVE

8.2.1 Data sources/Uses

The Smart Mobility Planner uses static data together with street information and the **real-time data** which are **provided by the 4 PTOs**. Although this was not initially planned, the project has now also added **three additional sources of data from different bike solution providers**: Villo (operating in Brussels), Libia Velo (operating in Namur) and Blue Bike (operating in Belgium). This has been done with a view to improving the customer services.

PTOs in Belgium have opened up real-time data in the past 2 to 3 years. Three PTOs involved in the SMP project (De Lijn, SNCB and STIB) have an open data platform, while for one PTO (i.e. TEC) opening up of data has been very recent. This data from TEC cannot be used as such, it requires the development of a certain reconciliation logic. Ultimately, SMP is intended to be **available as an open data source**.

8.2.2 Data flows

SMP is based on the **open source** route planning solution called **OpenTripPlanner** (OTP) which was developed in Portland by TriMet, the public transportation agency serving Portland, USA. Nextmoov injects **static data of the PTOs** into OTP together with **street information** and the **real-time feeds** of the 4 PTOs, which also existed previously, but not in a format compatible with the OTP.

The real-time data are sent by the PTO vehicles, but each PTO uses a different system. When it comes to trains, for example, they receive the information about the track on which the train is located, the station where it last stopped, etc. TEC and De Lijn on the other hand use the Global Positioning Systems. In order to use this data in OTP, Nextmoov has developed a **system of reconciliation between real-time information sent from the network and static data based on the scheduled departures**, which are then fed back into the route planner.

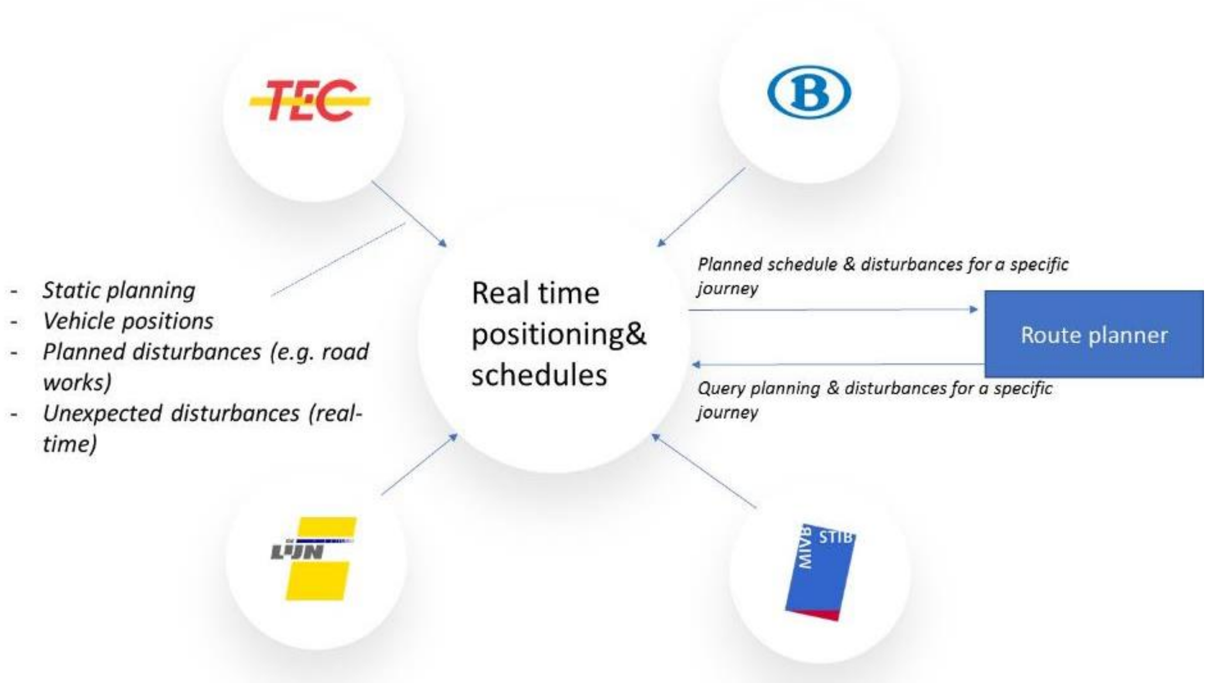

Figure 29 Overview of the reconciliation system of Nextmoov

8.2.3 Relevant big data policies

Belgium's consumers are less satisfied with the quality of both passenger rail transport and urban transport than the EU average.[112] Additionally, within the EU, the number of hours spent by commuters in road congestion in Belgium is also quiet high, making it the 3rd worst country

[112] See Page 28, DG MOVE (2019). Transport in the European Union: Current Trends and Issues', March 2019 available at https://ec.europa.eu/transport/sites/transport/files/2019-transport-in-the-eu-current-trends-and-issues.pdf

in the list after Malta and the United Kingdom.[113] In this context, some policies adopted by the Belgian Government have fostered new initiatives to optimize mobility and improve consumer experience.

The project on SMP has received funding through a specific budget provided by 2 Belgian ministries, The Ministry of Mobility and The Ministry of Digital Agenda, which launched a call on “Smart Mobility Belgium”[114]. The aim of the call is to facilitate public private partnerships and the use of open data to improve mobility.

The “Smart Mobilty Belgium” call builds upon the policy of using government held data for beneficial purposes. This policy has been promoted at the European level through the EU Directive 2013/37/EU and it has been also adopted in the federal legal system of Belgium.

As far as the terms used and the definitions given to them are concerned, Belgian law departs from the European text in two respects, but only terminologically. As regards the scope of application, Belgian law seems to be marginally narrower than what is envisaged under the Directive. Unlike the Directive, the Belgian text contains an exception which excludes the use of unfinished or incomplete administrative documents.

The Belgian law also envisages the creation of a “Public Sector Information” sectoral committee, which is tasked with giving prior authorisation for communicating documents containing personal data, and with delivering opinions on the “open data strategy” and anonymization methods used by public authorities. The EU Directive is also being transposed and used as regional law by separate regions in Belgium.

One of the key results of this process has been the opening up of data. A federal strategy of 'Open Data' was approved in July 2015, whereby it was suggested that public data belonging to the federal government should by definition be accessible, with a few exceptions based on privacy and security. In concrete terms, the Open Data strategy provided 14 key guidelines, which among others, include:

I. To freely re-use data without reference to the source as this makes it easy to combine data sets for the development of innovative applications;
II. To provide data in technical formats that facilitate reuse;
III. By 2020, the federal government should proactively make its data available and this is not something that should be done just on demand;
IV. To set-up a single federal portal with all publicly available and usable data;
V. To ensure maximum continuity: "re-users" must be able to rely on the availability of data in the future as well; and
VI. Each public service should develop an open data strategy and appoint a responsible person. The Open Data approach should cover, among others, aspects related to environment, mobility and economy.

[113] Ibid, Page 16.

[114] https://www.smartmobilitybelgium.be/pourquoi

Belgium has already opened some of its public data through its government open data portal (i.e. data.gov.be), where more than 473 datasets related to transport are being provided.

As part of the Open Data Strategy, the Mobility and Transport Federal Public Services has chosen to make available to the public a series of data from all its services. This data can be reused free of charge, in the spirit of the European Directive on the re-use of public sector information.

8.2.4 Main technological challenges and issues

When Nextmoov started scraping the **websites of the PTOs**, it realised that they were **poorly coded**. This, Nextmoov believes, may be attributable to the inefficiency of IT departments of big public transport agencies, who may underperform despite considerable budget available for development of new tools and solutions.

In terms of **infrastructure** Nextmoov relies on the hosting company called Digital Ocean based in the Netherlands for all their services. This **hosting company** operates a data center for Nextmoov and provides servers. Previously, Nextmoov tried cloud solution providers such as Azure, and Amazon Web Services. However, it chose Digital Ocean because it is based on similar business principles as Nextmoov, which is an important factor for the company in choosing all its tools, from communication to project management. It chose a European provider for multiple reasons. First, due to a perceived legal obligation to have the data hosted in Europe and due to physical proximity, which impacts the speed of the servers. Second, in order to support European know-how, for reasons related to cyber security. Thirdly, to contribute to the digital autonomy of the EU.

8.2.5 Big data dimensions and assessment

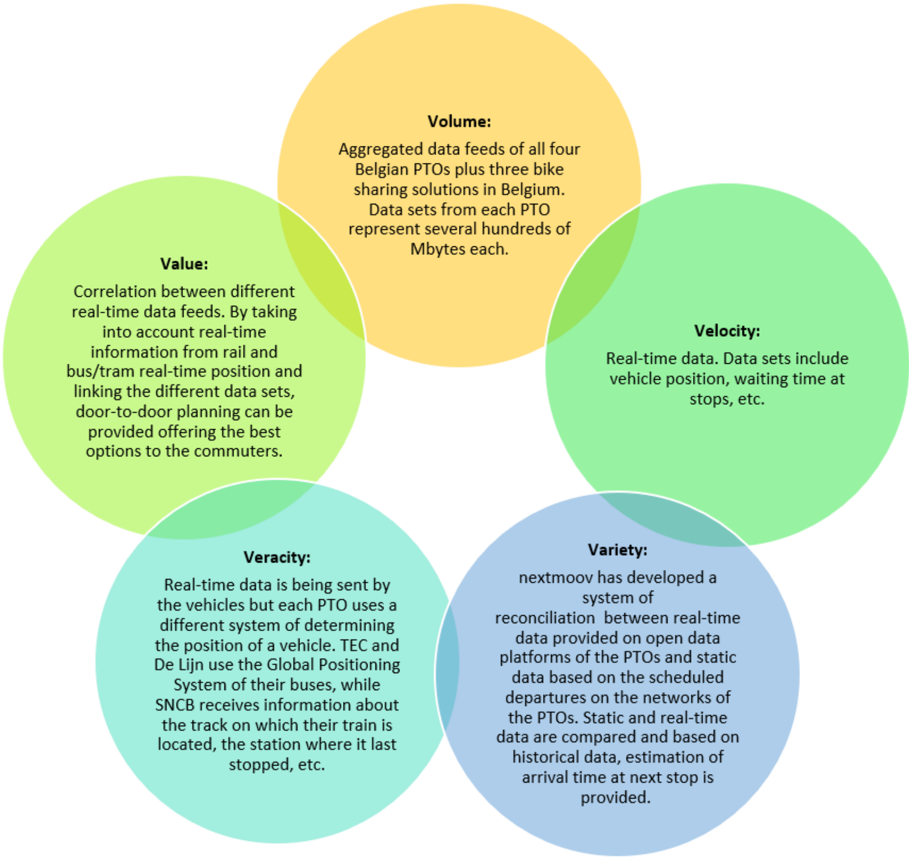

Figure 30 Big data dimensions and assessment in case study 6

8.3 BUSINESS MODEL

Nextmoov's business model is presented below.

Key partners	Key activities	Value proposition	Customer relationships	Customer segments
The customer segment i.e., the public transport organizations in Belgium are also key partners, as they are a source of data. Public agencies and research institutes, such as Vias Institute and Stratec.	Development of Innovative data-based mobility solutions (routing, e-ticketing, predictive analytics, etc.) for public and private clients in Belgium.	Smart mobility and routing solutions for Belgium, based on merging real-time datasets of the Belgian Public Transport Organizations.	Through the consortium structure involved in development and operation of "Smart Mobility Planner" application. Through web and mobile based platforms.	Public Transport Operators (TEC, STIB, De Lijn and SNCB) in Belgium as well as daily commuters.
	Key resources	**Internalization of costs and benefits**	**Channels**	
	Relevant open and closed datasets, hosting infrastructure; human resources; and increasing reliance on AI.	No external costs.	Electronic format	

Cost structure	Revenue streams
HR costs; costs for infrastructure used for storage and analysis of data; costs of infrastructure maintenance.	Public tenders and private clients. The development of the application being discussed in the case study has been paid public money. The route planner is available free of charge. It will be used by next moov, Stoom link to provide new services either directly to B2B market or B2C. The business model has still to be defined but is likely to based on licenses for B2B and % of ticket sold for B2C.

Figure 31 Business model canvas of Nextmoov

8.3.1 Business processes

SMP won the **Smart Mobility Belgium public call for tenders** in July 2018 with a total budget of EUR 400,000. The consortium responsible to develop SMP is composed of Nextmoov and the four Belgian PTOs mentioned in the section before.

As mentioned before, in addition to developing SMP, Nextmoov also works on other projects within the framework of the joint venture **Stoomlink** together with the Vias Institute and Stratec, focusing on **mobility data and front-end for end-customers**. In terms of big data expertise, Nextmoov collaborates with experts from the Vias Institute.

In the future, Nextmoov plans to continue submitting tenders to the calls of public transport authorities in Belgium, while also developing their portfolio of private clients.

8.3.2 Relation to big data initiatives

There are several different initiatives in Belgium that focus on optimizing mobility and improving consumer experience by using Big Data. Besides the initiatives of PTOs, which themselves provide open data (see for instance last initiative of the rail network operator Infrabel), a number of private companies are also developing mobility companion applications for cities like Brussels (e.g. Joyn Joyn) and Antwerp (e.g. Pikaway) These applications collect different types of mobility data (i.e. number of available shared cars, location of bus/tram stops etc.) and provide route planning and ticketing services to their customers.

The SMP developed by Nextmoov complements these and other initiatives that have been launched in Belgium pursuant to the open data and smart mobility policies of the Belgian Government, which are discussed in Section 10 below. A detailed list of various projects being implemented within the framework of these policies can be found on the following links:

- https://www.smartmobilitybelgium.be/projets
- https://data.gov.be/fr/search/apps?f%5B0%5D=im_field_category%3A36

8.3.3 Illustrative user stories

SMP has been developed as a solution to a situation where PTOs in Belgium have been using their own distinct algorithms of route calculation to support their respective mobility applications. In practice this has meant that results for the same route, with the same departure time and destination in Belgium, would be different in each PTO application. The reason behind this discrepancy is that the separate application of each PTO takes into consideration data only from its own public transport network, while neglecting data from other public transport networks, i.e. the bus networks do not consider the data from rail network of SNCB and vice versa. Additionally, TEC, De Lijn and STIB do not consider the real-time data from their networks for the route planning aspects. They only display real-time data for departures at the stations, but this data is not taken into account for route planning. Instead, they use static data without taking into account any information related to delays on their network and as a result the performance of the route planning application of these PTOs is rather low. The **SMP project intends to address this problem by covering entire Belgium and using the same algorithm for the four involved PTOs by merging the four feeds of real-time data.**

The initial objective of the project was limited to developing only a web version of SMP, but eventually it was decided to develop a **responsive mobile application**. Nextmoov and the consortium working on SMP also decided that the application would reflect **different public transport options for the users**. The first option would provide an **optimised route to an individual user** – where optimized routing results would be displayed, taking into account information such as weather, if a user owns a bike or not, if a user has a monthly pass for a certain PTO network or not, etc. (these are some parameter that applications such as Google Maps do not take into account). The next option would indicate **the fastest route** for users who intend to reach their destination as soon as possible. The third option would highlight the **eco-friendliest** solution, combining bike, rented bike and walking for small distances OR train and bike for longer distances. In the future, the **solution could also integrate ticket payments**, based on the technical expertise from another project, called Modalizy, which is implementing e-ticketing and where Nextmoov has been involved.

8.4 ANALYSIS OF SOCIETAL ASPECTS

8.4.1 Economic and political

Issues related to the political framework in Belgium

Fragmentation of PTOs and regional fragmentation in Belgium adds to the complexity, when it comes to – collaboration between PTOs, sharing objectives and implementing coordinated solutions for the users.

Lately, however, a shift towards identifying shared objectives has been observed. This may have resulted on account of common threats posed by the big players such as Google and Uber.

Inefficiency of the PTOs in terms of providing innovative routing solutions

The perceived inefficiency of PTOs has been mentioned earlier. In the view of Nextmoov, a possible reason for this could be the structure of public agencies in general and a lack of awareness about the objectives among employees. In addition, outdated equipment, rigid schedules, limited availability of technical tools, inefficient and frequent meetings, etc. can also be viewed as contributing factors. The challenge for PTOs is to compete with new and more agile players by becoming themselves more agile and/ or by collaborating with new players such as Nextmoov.

Before SMP, no mobile solution was available in any of the four PTOs. The user had to scroll down the PDF document with a timetable to know about departure and arrival information. Even today the transport providers in Belgium are slow in adapting to innovation and their mobile applications are only updated once or twice per year. In contrast, CityMapper, Messenger and other similar big mobile application are updated every two weeks. But, owing to the reasons mentioned in the paragraph before, it is difficult for the PTOs to keep up with the pace observed in the private sector.

Cooperation with PTOs

On the project level, in order to ensure efficiency and leverage the involvement of the PTOs, the project is managed using a **two-level governance model**. Every month, there is a meeting of a Steering Committee with Marketing Directors and Innovation Managers of the companies. Team Meetings take place on a weekly basis, bringing together developers and graphic designers from Nextmoov and PTOs. This allows for the flow of ideas from the PTOs, but also ensures that the objectives of the project are met.

Issues related to digital autonomy

Google Maps poses important questions for the PTOs in terms of digital autonomy. In Brussels a trend referred to as "**mobility wars**" can be observed, as there are numerous providers of the same service. An example of this trend can be seen in the proliferation of e-scooter service providers in the city. In this context, Google Maps may have a lot of influence to shape and decide what kind of transport services are featured in a certain area. The PTOs are also aware of this phenomenon and it has been observed in case of Uber, where their services were favoured by Google Maps and reflected as the fastest routing option. Now, a similar tendency can be observed with Lime (an e-scooter sharing service that originated in California)[115], which seems to be favoured by Google Maps over Troty (an e-scooter sharing service developed by a Brussels based start-up in 2018). This threat translates into an important incentive for the PTOs to try to offer a comparable tool over which they could have more control, contributing thereby to national and EU digital autonomy over traffic management, road safety, smart cities and critical infrastructure (physical and IT).

8.4.2 Social and ethical

Public perception and lack of trust

When it comes to data availability, one particular type of data which is considered very valuable is **data related to the usage of the Public Transport network** gathered via passes (such as records of information every time someone enters or exits the network). PTOs in Belgium are not using this data as much as they could, hence its potential has not been exploited to the full extent. This is something that Nextmoov desires to be implemented in the future. A good example of this is "TFL" in London where data about entries and exits on subway network is available as open data.

The main reason for the unavailability of this data is suggested to be **privacy issues**. But Nextmoov also perceive the **lack of comprehension by the public** as an important factor.

A recent example of such public misperception was observed during the introduction of a MOBIB card[116] in Belgium, when the public expressed concerns about being tracked and monitored. Nextmoov also received some similar feedback regarding the SMP when some individuals testing the application inquired whether they could be tracked through their phones during the test phase and if their position could be identified by the application. In the opinion of Nextmoov such concerns are understandable considering the privacy issues being discussed in relation to Google and Facebook, but the lack of public comprehension is also very high. This

[115] It was first launched in Brussels and Antwerp in 2018.

[116] The Mobib Card allows public transport users to combine all their SNCB, TEC, STIB and De Lijn season tickets in a single card.

is moreover surprising considering that people often tend to neglect privacy policies and cookie banners when they want to use applications or any sort of information system that makes their life easier. It shows that privacy remains a concern. A possible way to explain this concern could be to consider that Nextmoov is a small company and the level of trust by the public is not as strong as it is in case of global players (in the future, the inverse could be true: the level of trust towards smaller and EU based companies may become higher compared to larger and often US based companies)

Nextmoov believes that gathering aggregated data could also be problematic in terms of possible **traceability of individuals**, for example in rural areas where the number of public transport users is low. A possible solution to address this could be to draw inspiration from the practice adopted by Uber, which offers a service called “Uber Movement” that visualizes how many Uber services are being used in different cities. However, this visualization is possible only when activity in a city surpasses a certain threshold. Data from cities below this threshold is not made openly accessible. In case of Brussels, adopting a similar practice of thresholds and making data accessible during a certain timeslot (such as peak hours) when the number of public transport users is high, could greatly contribute towards improving mobility from the technological standpoint, as opposed to not opening this data at all. The **information during peaks hours could help improve the traffic flow** and consequently reduce related costs, such as resources related to increased security, etc.

Nextmoov is currently discussing with PTOs to push the discussion in this direction and it has also been planning a campaign where to generate SNCB tickets at discounted prices for off-peak hours in their Modalizy[117] application to further capitalize on this.

The Table below lists the societal and ethical issues in businesses of Nextmoov according to the six topics identified by (Debussche, César, Hong, Nordeng, & Waldenfels, 2018): trust, surveillance, privacy, free will, personal data ownership, discrimination and environment.

Table 27 Overview of societal and ethical aspects and relevance to Nextmoov

Aspect	Description of aspect	Relevance of issue to SMP and Nextmoov
Trust	Increase of decrease in trust in by or between the data users, data facilitators and data suppliers.	Level of trust expressed by public in SMP and Nextmoov is not high, since compared to bigger global companies Nextmoov is relatively a small company. Proposed Solution: There is a need to better inform public. Opportunity: companies should create value from their modest size and geographical location (local, regional, national, European).

[117] Modalizy provides a comprehensive solution to the challenges of business mobility and management of business travel expenses by facilitating the processing of business travel expenses (paying for a ticket, collecting proof of payment, submitting bills of costs, etc.).

Surveillance	The ability to collect large amounts of data enables the data user to better understand processes in society.	Gathering aggregated data could be problematic in terms of possible traceability of individuals, especially in rural areas, where public transport users are relatively low. Proposed solution: Introducing a threshold system, whereby data for cities below a certain activity level threshold is not made public.
Privacy	The involvement of privacy sensitive information	The data related to the usage of Public Transport network, gathered via passes, is considered valuable. But this data is also deemed sensitive in view of privacy concerns. The concerns related to the sensitivity of this data can however be attributed to the misperception of public and raising awareness in public could help address this issue.
Free will	Big data analysis and associated predictive analysis that threatens human autonomy	The use and analysis of open data in this case study does help shape the travel choices that commuters may make, however, it cannot be said that this would impact the free will of the commuters.
Personal data ownership	Confusion of who owns the data or loss of data ownership by data suppliers	There is no concern related to personal data ownership, as the data used in this case study is open data.
Discrimination	Discrimination of population groups based on generalisations derived from big data	There is not concern related to discrimination in this case study.

8.4.3 Legal

In terms of legal matters, Nextmoov has not faced any particular issues. They do not have in-house legal expertise, so they rely on law firms and legal consultants, specializing among others in open source projects, whenever a legal matter needs to be addressed. In the particular case of SMP they have also consulted legal teams of the 4 PTOs.

As regards legal aspects, the main focus for the company was on privacy and personal data protection, which required the largest part of legal-related work. Before and after the entry into application of the General Data Protection Regulation (**GDPR**), a huge difference is noticed. The introduction of the new legal framework is considered a positive evolution for the end-users, however it has substantially complicated the company's work. A few concepts are considered particularly interesting and positive, such as the requirement to provide information to the data subjects and the requirement for organisations to document everything they do in terms of personal data processing.

Certain other requirements have purely complicated things without bringing actual benefits, such as the cookie requirements, which do not arise from the GDPR but from the e-Privacy legislation at EU level. There is a very negative sentiment towards these cookie banners and opt-ins, which are a nightmare from a user experience standpoint.

Another issue seen or feared is that big players will be able to find ways to circumvent certain of these obligations, whereas small players cannot as they do not have the money to figure these things out. Additionally, small players cannot take the risk to infringe any of these legal requirements that may lead to fines and claims for damages from individuals.

The company often relies on open data to develop the SMP, this is possible since PTOs are now really opening their data, particularly real-time data, which constitutes a big change with a few years ago. While certain PTOs have established open data platforms, a lot remains to be done in Belgium in terms of opening up PSI data. In other projects they have worked together with private companies which have shared their closed data. This was always arranged through data sharing agreements, which tend to be pretty clear. While the companies still consider it would be easier if all data were open, that is considered more a business matter (i.e. negotiating an appropriate license) than a legal one.

The issue concerning digital autonomy was already mentioned above. In this respect, another interesting finding, with potential consequences from a competition law perspective, is the fact that one of the e-scooter companies, Lime, refuses to share its data. Lime has signed an agreement with Google Maps, giving Google access to the location data of its scooters. Nextmoov has experienced that it is not possible for them to build an application, in the same way as GoogleMaps, and add Lime for instance as an eco-friendly transport option for the last mile of a trip since Lime refuses to share its data with other players. Nextmoov asked the Ministry of Mobility in Brussels to require from any **provider operating a shared mobility solution to open up their data** which is the case in some cities, such as Seattle for example. Otherwise, Nextmoov fears a potential **threat of the formation of monopolies,** as it observes based on the link of Google Maps with Uber and Lime. This leads to questions from a competition law perspective, considering Google's market power and the way it is using this power.

The table below provides an overview of the legal aspects in relation to big data encountered by Nextmoov and the findings in this respect.

Table 28 Overview of legal aspects and relevance to Nextmoov

Aspect	Description of aspect	Relevance of aspect to SMP and Nextmoov
Privacy and data protection	Consideration and implementation of the requirements of the GDPR and other types of privacy and data protection	Pre- and post-GDPR, a huge difference is noticed. The introduction of the new legal framework is considered a positive evolution for the end-users, despite having substantially complicated the company's work. A few concepts are considered particularly interesting and positive, such as the obligation to inform data subjects and the accountability requirements. Other requirements have purely complicated things without bringing actual benefits, in particular the cookie requirements, which elicit a largely negative sentiment in the organisation. Cookie banners and

		opt-ins are considered a nightmare from a user experience standpoint.
(Cyber-)Security	Awareness and implementation of certain general or specific (cyber-)security requirements	The organisation notices an overall very low level of security of many big companies. It considers the use of up to date and new technology an easy way to address security threats and minimize issues in this respect. The company moreover considers the use of open source technology a good solution for cybersecurity issues. Open source software gets updated really quickly, including in case a cybersecurity issue is detected. The huge open source community can fix this type of issue straight away, whereas a security issue with proprietary software of a tech giant can only be fixed by the competent person at that organisation, which can take a lot longer and therefore entails a much larger risk. The company also believes that since it relies on a European provider for its infrastructure, its cyber security is further reinforced, as such infrastructure is less vulnerable to interference/attacks from outside Europe.
Anonymization and pseudonymisation	The use of any type of anonymisation or pseudonymisation technique, i.e. techniques to ensure personal data can no longer be attributed to individuals without using additional information or to reduce the identifiability of individuals	The company takes a very pragmatic approach, building gates and other technical measures into its applications to ensure that unwanted identification cannot and will not happen, thereby using security measures and pseudonymisation techniques as a means to comply with data protection law. It is considered much easier to build this type of technological barrier than to engage in the specific processing activity and being required to comply with all legal requirements presented by privacy and data protection law. In this respect, the company considers the approach taken by Uber Movement, not opening the data when activity in a certain area is below a specific threshold, a good solution and in any case much smarter than simply not opening up any data, as this does not help improve mobility.
Free flow of data	Awareness of any data localisation requirements hindering the free flow of non-personal data for the organisation	The company's servers are located in Europe. One of the reasons for this is the idea of the organisation that there is a legal obligation to have the data hosted in the EU. They were under the impression that while you were allowed to rely on a non-EU company for providing the hosting service, such as Amazon and Microsoft, the data does need to be hosted in Europe.
Intellectual property	Protection by the organisation of its service/product through intellectual property rights and hindrance for the	The company relies a lot on open source technology. The application being developed by Nextmoov will be open source. Therefore, in the future it will be possible for citizens to use real-time open data.

	organisation of third parties claiming or otherwise asserting intellectual property rights	
Open data	Awareness and use of open public sector information	The company often relies on open data. This is possible since today, public transport operators (PTOs) are really opening their data, particularly real-time data, which constitutes a big change with a few years ago. One PTO opened up its real-time data only very recently. The three other PTOs however have established open data platforms where you can get the real-time data. Despite this progress, the company feels that a lot remains to be done in terms of opening up PSI data.
Data sharing agreements	Use of data from non-public organisations and the organisation of the data sharing	When working with private companies, those companies share their data with Nextmoov through data sharing agreements. These agreements tend to be clear. The issue in this respect is considered more a business matter than a legal one, i.e. it's a quesiton of negotiating an appropriate and acceptable license.
Competition	Issues related to unfair competition, particularly related to agreements between undertakings, abuse of dominance and mergers and acquisitions	It is feared that big players will be able to find ways to circumvent certain of the legal obligations, whereas small players such as Nextmoov cannot as they do not have the funds to figure these things out.

8.4.4 Environmental

Though no environmental issues have been identified in the specific example of SMP, it will not be wrong to say that optimization of mobility can have a net positive impact on environment, if it can help reduce road congestion and primary reliance on cars for movement in Belgium.

However, route planning applications will have to tackle the challenge of creating bottlenecks by guiding all commuters to the same optimal path. If this is not addressed then it may lead to a rebound effect, contributing to further congestion instead of addressing it.

Such pitfalls can be avoided through real-time tracking of traffic flows and by creating limits that prevent the same path from being suggested to commuters after a threshold is reached.

8.5 ASPECTS RELATED TO OPPORTUNITIES, BARRIERS AND LIMITATIONS

8.5.1 Opportunities

Generating people's movement could open up the entry for Nextmoov into **new services** aside from transport. The current development allows disposing of an integrated data set across different operators that is also open for integration of other data sources. This is a step forward in Belgium. The work of Nextmoov has also allowed cooperation of different Regions around a common product to solve an important mobility issue. Nextmoov's expertise could be beneficial for other sectors and industries which face similar challenges in terms of the structural inefficiency. The same agile approach and work culture could be brought into other fields, for example in finance, to improve the customer experience. By combining data, application solutions and expertise in mobility and road safety, Nextmoov and Stoomlink are able to propose a range of solutions to different markets, starting from addressal of local needs to providing complete IT solutions.

In the future, Nextmoov plans to **incorporate predictive analytics**. This concept is depicted in the figure below.

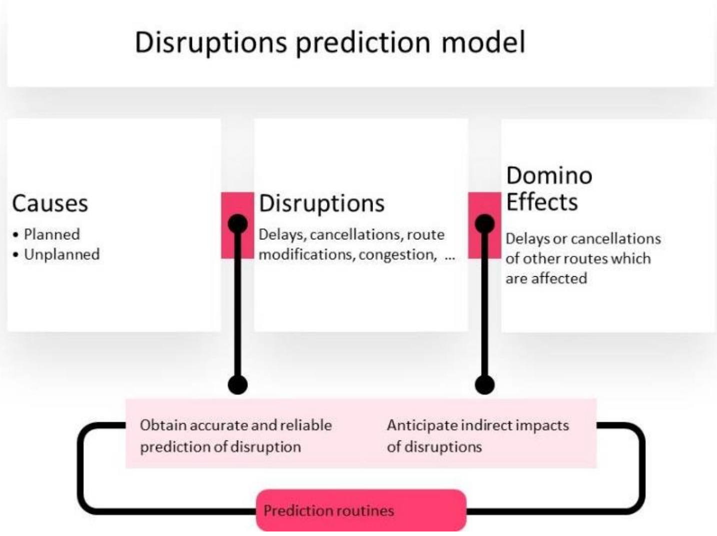

Figure 32 Concept of predictive analytics in Nextmoov

This requires a lot of historical data and Nextmoov will start to systematically save data feeds related to delays from unified real-time data (at the moment, no cross-operator historical data exists) and then extract patterns together with big data experts from VIAS and Stratec. This work is expected to start in 2020.

8.5.2 Barriers and limitations

- **Lack of Vehicle Data Access:** Traffic data generated by GPS in cars or more generally specific sets of data stemming from vehicles could prove useful to mobility application developers such as Nextmoov, for facilitating further innovation. Currently, such data is completely closed and very expensive to acquire.

- **Lack of Realtime/Dynamic Data:** At present, latency also poses a challenge. The data available from PTOs is not real time and as much dynamic information as expected is not available.
- **Application Fatigue for Customer's:** 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 "one stop shop" application can be off-putting and prevent expansion of userbase
- **Digital Autonomy:**
    - *Operational Perspective*: Large service providers (such as Google) have capabilities that can allow them to undermine digital autonomy from an operational perspective (e.g. traffic management, road safety, smart cities). On one hand these capabilities can be used to nudge consumers to use mobility services that may be owned by or pay the service provider. On the other they can be used to influence traffic flows resulting in health as well as environmental hazard. The latter was recently observed in Belgium where a GPS Navigation application owned by Google "Waze" was criticised by various municipalities around Brussels for re-directing traffic away from the main road and into such municipalities with residential zones, thus increasing the health and safety hazards.
    - *Infrastructure Perspective*: There are also concerns about data storage service providers from outside Europe attracting application builders from Europe. Such storage service providers offer attractive deals, by also providing technical support on data analysis, but can compromise digital autonomy, if vast amounts of data ends up being stored with service providers from outside Europe.

8.6 SOURCES OF INFORMATION

8.6.1 Interview

An interview with the CEO of Nextmoov was conducted.

8.6.2 References

Deloitte (2019). A new deal for mobility in Belgium

Directorate-General for Mobility and Transport (DG MOVE) of the European Commission (2019).Transport in the European Union: Current Trends and Issues', March 2019

Smart Mobility Belgium - https://www.smartmobilitybelgium.be/pourquoi

DG MOVE (2019). Transport in the European Union: Current Trends and Issues', March 2019 available at https://ec.europa.eu/transport/sites/transport/files/2019-transport-in-the-eu-current-trends-and-issues.pdf

Deloitte (2019). A new deal for mobility in Belgium

9 Case study 7: “Big data and intelligent transport systems”

The case study on “Big data and intelligent transport systems” explores the eco-system of on-demand mobility known to be heavily reliant on big data, with example of the mobility on-demand service provider ioki, Germany. Interviews were conducted with Smart City of Deutsche Bahn (DB) and ioki.

9.1 CASE STUDY OVERVIEW

9.1.1 Organisation/Stakeholders

Mobility on-demand (MOD) is an often-used term to describe new and at short notice available mobility offers which include the use of digital services. It serves as an umbrella term for all mobility services that are available to the user catering specifically (or very closely to) individual requirements, such as departure point, time of departure and destination.

In this term, the included mobility offers are:

- conventional taxi services, which represent the classic and first form of ride hailing services;
- services operating in areas with low demand like in rural areas, as a substitute for scheduled services, such as the call bus systems;
- “interface” services between car ownership and the use of public transport, such as car sharing;
- services that offer empty spaces on private journeys using Internet-based booking, such as the ridesharing systems;
- ride-pooling services, which combine passenger requests and then serve them using a calculating algorithm;
- offers that are limited to providing information on the possible travel options for the customer and are not physically active transport service providers. Mobility-as-a-service (MaaS) integrates the available means of transport and provides the customer with an individually adapted travel option.

In this case study, we take a general look at MOD and examine various aspects of it. In addition, we have examined the on-demand mobility service provider ioki[118].

ioki started as part of Smart City | DB, which is part of Deutsche Bahn AG (DB) and deals with the design of new services around railway stations. Smart City is also closely linked to the topics of urban mobility, logistics and urban habitat development.

Smart City | DB (hereafter “Smart City") is an internal group program (founded in 2017) of Deutsche Bahn AG (DB). Smart City develops new mobility, infrastructure and logistics concepts for the “first and last mile”. After a successful proof-of-concept of a new business model, the concepts are returned for implementation in the respective business areas of Deutsche Bahn (e.g. DB S&S). Smart City carries out data-driven analyses in each new potential business case.

---

[118] ioki, “ioki - Inspiring smart mobility,” 2019. [Online]. Available: https://ioki.com/. [Accessed: 05-Apr-2019].

Here, the special focus is on the needs that arise in urban areas. For this purpose, the urban analytics team plays a crucial role within Smart City.

Smart City consists of around 20 employees. The urban analytics team with 5 employees carries out the data-driven analyses (current April 2019).

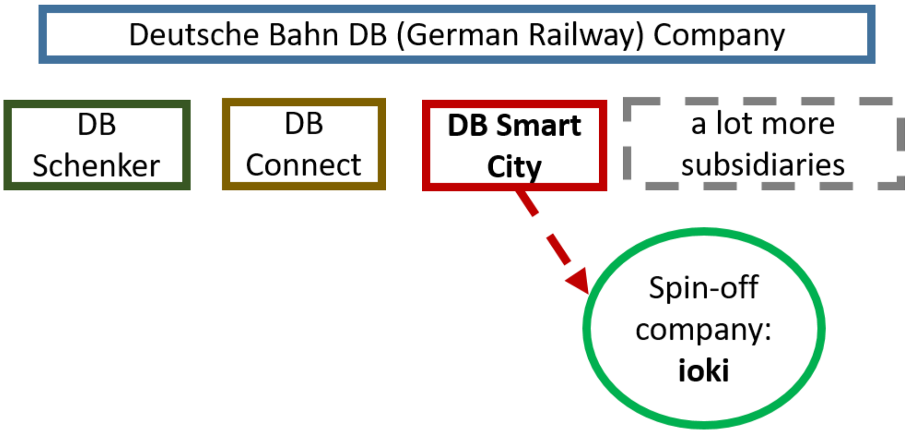

Figure 33 Relation between German Railway (DB AG) and ioki

Meanwhile, ioki has been spun off as an independent subsidiary company that implements on-demand mobility in various German cities. ioki and Smart City are working in deep partnership and do various projects together.

ioki is a spin-off company of DB founded in 2016 and offers mobility analytics, on-demand mobility services and autonomous driving services. It started as an internal pilot project in Frankfurt am Main, which is still running. The company works together with public transport operators and integrates their services into the public transport systems, for example to improve mobility offers in rural areas. Ioki currently employs 75-80 people.

The aim of ioki is to offer on-demand mobility integrated into public transport, which is operated autonomously in perspective. The two components of the service, first on-demand mobility, second autonomous driving services are currently developed separately. They should be combined when the development will have progressed enough to do so.

Ioki's services cover **on-demand digital platform**, on which their **on-demand mobility services** can be based on. Simultaneously they are *developing* **autonomous driving services**, like selected routes, to make them market-ready, and combine them with their other services. As a last building block, they offer **mobility analytics** for the cooperation partners like public transport operators.

Figure 34 gives a schematic overview of the involved parties and stakeholders in MOD. MOD is often referred to in the context of **urban and multimodal** mobility, as new mobility offers are often initially implemented in larger cities or conurbations with greater demand. MOD mobility offers are not limited to a certain mode, but often combine different modes in such a way that the customer's mobility request is implemented as efficiently/quickly/cost-effectively as possible.

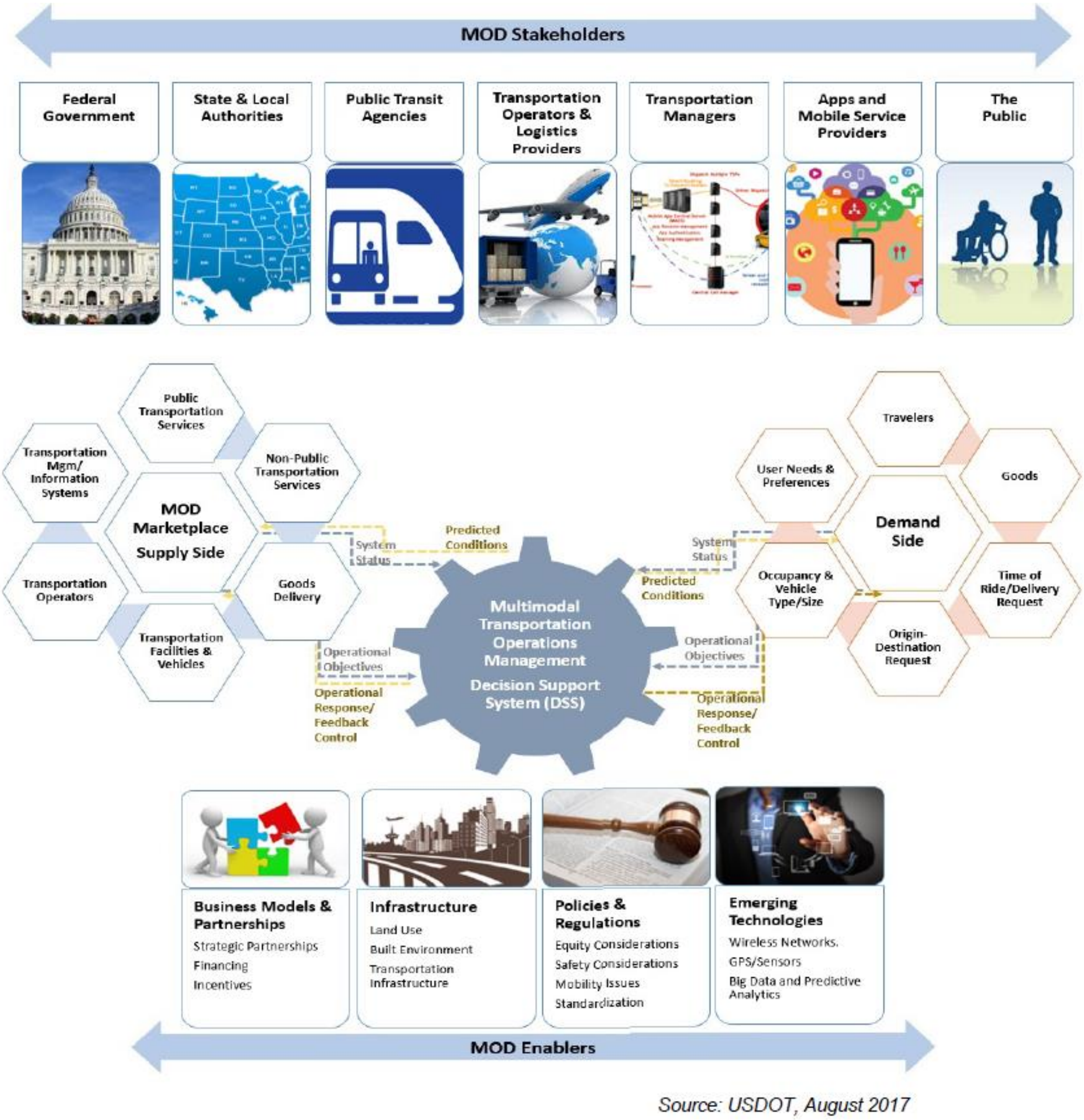

Figure 34 On-demand mobility Ecosystem

Various stakeholders are involved in the context of MOD. These stakeholders have different roles and needs in the conception, implementation, use or execution of the mobility offers. An overview of the individual stakeholder groups can be found in the table below.

Table 29 MOD involved stakeholder

Stakeholder	Stakes
Governmental institutions	These organisations define transportation strategies, policies, and legislation. They implement planned strategies, invest in pilot programmes and formulate guidelines for national development.
On-demand mobility service provider	Want to implement new business models, optimise new services as barrier-free as possible and adapt it to the given environment. They influence the

	MOD ecosystem through their services, but they are also influenced by their environment and the framework conditions.
Cities, communities	Want to ensure smooth mobility flows, save infrastructure costs, implement urban planning and strategies, improve and maintain urban space; must deal with limited budgets and are subject to strict guidelines. They also do the strategic urban and traffic planning and are responsible for the local infrastructure.
Public transport operators	Want to make public transport attractive, cover their costs, dependent on capacity.
Citizens	Wants a noise- and emission-free living space and affordable, flexible transport services, commuters need reliable mobility services.
Manufacturers / Apps and Mobile Service Providers	Want to bring new or existing vehicles and technologies on the market. Enable the implementation of on-demand services through mobile ticketing, payment and navigation services.
Customers	Want affordable and flexible mobility offers, want their personal data and travel information to remain protected. Business customers want an business friendly not to regulative environment.

9.1.2 Sector/Mode

On-demand mobility is one of the most important trends in transportation and in the global economy today. On-demand mobility (moving people, goods or services) is based on a mobile application that allows for short-term scheduling and uncomplicated payment.

9.1.3 Case study motto

Mobility on-demand services offer many new opportunities as well as challenges in the digital age.

9.1.4 Executive summary

In this case study we explore the eco-system of on-demand mobility, examine the different business processes and models and compare it to the concrete example of the German MOD service provider ioki.

In the second part we investigate the technological background and highlight the most important challenges yet to overcome. We explore societal aspects based on literature research and the information from our case study partners of ioki and Smart City.

9.2 TECHNICAL PERSPECTIVE

9.2.1 Data sources/Uses

In the below *Table 30*, data sources are listed, which concern urban or connected mobility and are typically used to implement or operate MOD services[119][120].

Table 30 Data sources with possible use for MOD services

Type of Data Source	Description of Data Collection Process
Traffic Flow Sensors	There are two categories of traffic sensors. First one is in-roadway sensors that are embedded in pavement or attached to road surfaces, including inductive-loop detectors and magnetometers sensors. The second category is over-roadway sensors that are mounted above the surface, including video image processor, microwave radar sensors, laser radar sensors, and ultrasonic/acoustic/passive infrared sensors.
Video Image Processor	A video image processor system typically consists of one or more cameras for taking vehicles' images and videos, a microprocessor for storing and processing images, and a software to apply computer vision and image processing algorithms for analyzing the raw image data and exploiting traffic parameters used in the traffic management [6].
People and Vehicle Generated	In this case positioning systems record spatiotemporal information of floating sensors while moving in the network. Examples for such technologies are Global Positioning Systems (GPS), mobile phone cellular networks, and Bluetooth. Devices equipped with these technologies (e.g., drivers' smartphones) can track vehicles or people's locations indexed in time to create their trajectory data.
Location-Based Social Networks	The Location-Based Social Network (LBSN) that generates users' spatiotemporal information is the most useful type of social media in transportation-related applications. The LBSNs are categorized into three groups: Geo-tagged-media-based, point-location-based, and trajectory-based. Examples for such social services are Foursquare, Yelp, Twitter, Panoramio and Flicker.
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. The main application of the smart cards is to collect revenue. The APC has been

[119] S. Dabiri and K. Heaslip, "Transport-domain applications of widely used data sources in the smart transportation: A survey," pp. 1–53, 2018.

[120] L. A. Klein, "Traffic Detector Handbook:Third Edition—Volume II," *US Department of transportation*, 2006.

	particularly designed for counting passengers in and out of public transport modes (e.g., bus and subway).
Environmental	Meteorological data consists of atmospheric information like humidity, temperature, barometer pressure, wind speed, evaporation, precipitation, snowfall, types of weather, etc. provided by public websites and agencies. Air quality data indicate the concentration of various pollutants in the air such as carbon monoxide, lead, ground-level ozone, nitrogen dioxide, sulphur dioxide, and particulate matter.

For the analysis of new potential places of use, we talked with our interview partners:

The identification of potential routes and assessment of a successful implementation of ioki is carried out by analyses with the following data sources:

- Geo-data: Map data, open source maps, information on social milieus, mobility data, and cartographic points of special interest.
- DB traffic model: The traffic model is an agent model that combines the traffic flows collected by the railways with purchased data from mobile phone providers. DB does not sell its traffic model data or offer advice to cities, but only uses it within the company.
- Open Data Portal of DB[121]. In this portal, the individual business units of DB provide selected data. Open data from cities, municipalities or the Federal Statistical Office are also used.

Amongst others the following technologies are used:

- a geo-information system
- software for traffic models (vissim)
- self-developed algorithms especially in JAVA and C++. The experience of the DB group in IT is helpful and most of the software is implemented internally.

As an example: To start operations in Hamburg, ioki analysed the area of interest first, following an agreement with the local transport provider Verkehrsbetriebe Hamburg-Holstein GmbH (VHH). Important parameters were, for example, the degree of motorisation of the population, the public transport frequency and the population density. The developed area lies more in the rural area around Hamburg and is only inadequately served by existing public transport.

Ioki also collects real time operational data, such as route history using GPS data and user booking data. The tablet installed in the vehicle sends this data to the Ioki servers via the Ioki platform. In the future, supply and demand simulations will be implemented that will consider external data such as event and weather data. Comprehensive predictive maintenance is planned for an autonomous fleet.

[121] "Open Data Portal DB." [Online]. Available: https://data.deutschebahn.com/. [Accessed: 05-Apr-2019].

9.2.2 Data flows

In the context of MOD, the data flow is a very complex and can be identified as a state-of-the art IoT infrastructure consisting of multiple interconnected layers[122][123]. The exact implementation of such infrastructure can vary based on the city population, traffic intensity, infrastructure development plans and existing transportation and many more.

In the case of on-demand mobility like ioki, there are exemplary data feed communications described in detail by Masek et al. and depicted on *Figure 35*[124].

A typical Traffic Management System (TMS) consists of four complementary phases, as depicted in *Figure 35*. In the first ***Data Sensing and Gathering*** phase, data is collected from various sources such as heterogeneous road monitoring equipment[125] that measures the important traffic-related parameters (e.g., traffic volume; speed; and occupancy of the road segments) over certain time intervals.

Next the measured data is forwarded to the traffic management entity (TME); the detected events are immediately reported over the deployed wireless communications networks (e.g., cellular systems). Then, the obtained data feeds are processed (aggregated) in the second phase known as the ***Data Fusion***, processing, and aggregation (DFPA). In the case of ioki additional data is collected from the devices mounted in the vehicles and the driver and traveller applications. The third phase, named the ***Data Exploration*** (DE), processes the knowledge from DFPA and computes the

- optimal routes;
- short-term traffic forecast;
- supplementary road traffic statistics.

In the case of ioki this data is merged with historical and other external data sources to obtain more precise traffic forecasts and plans that can be than returned to the drivers. The final phase, termed the ***Service Delivery*** (SD), distributes the resultant knowledge to the end users/commuters (e.g., drivers, private companies, emergency services, etc.).

[122] R. Khan, S. U. Khan, R. Zaheer, and S. Khan, "Future Internet: The Internet of Things Architecture, Possible Applications and Key Challenges," in *2012 10th International Conference on Frontiers of Information Technology*, 2012, pp. 257–260.

[123] R. Petrolo, V. Loscrì, and N. Mitton, "Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms," *Trans. Emerg. Telecommun. Technol.*, vol. 28, no. 1, p. e2931, Jan. 2017.

[124] P. Masek *et al.*, "A Harmonized Perspective on Transportation Management in Smart Cities: The Novel IoT-Driven Environment for Road Traffic Modeling," in *Sensors*, 2016.

[125] B. Chen and H. H Cheng, *A Review of the Applications of Agent Technology in Traffic and Transportation Systems*, vol. 11. 2010.

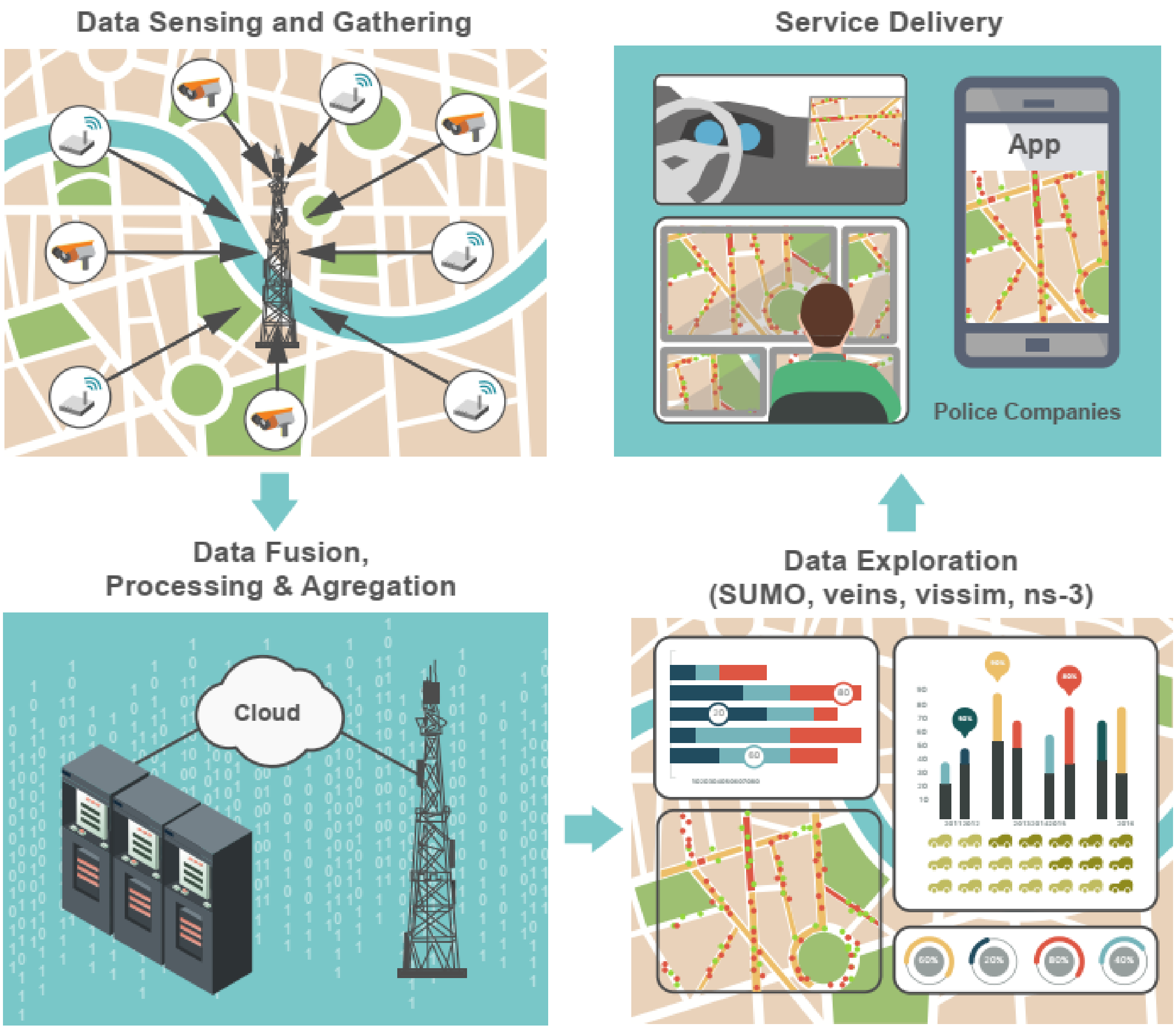

Figure 35 Communications chain of data feeds in smart transportation (parts of the traffic management system)

9.2.3 Main technological challenges and issues

With the growing data sizes, increasing variety of data and the increasing velocity with which the data is arriving, one major data challenge turns to be the effective data collection and processing that leads to improved service results. Data fusion (DF) is a collection of techniques by which information from multiple sources are combined to reach a better inference[126].

The purpose of DF is to produce an improved model or estimate of a system from a set of independent data sources. For traffic applications, the desired model is the state vector of the traffic phenomenon. These estimates may include current or future vehicular speeds, travel time, vehicle classification and similar topics of interest to travellers and traffic operators.

[126] N. E. El Faouzi, H. Leung, and A. Kurian, "Data fusion in intelligent transportation systems: Progress and challenges - A survey," *Inf. Fusion*, vol. 12, no. 1, pp. 4–10, 2011.

However, the challenge is in building and applying the right Big Data technologies, Machine Learning, Deep Learning and Artificial Intelligence algorithms to create accurate real-time model[127].

Transport modelling is important and essential for estimating travel demand and offering valuable information to policy makers and transport planners[128]. The modelling process that has evolved from static to dynamic capturing travel behaviour in terms of time-dependent conditions and information, and from an aggregate to a disaggregate representation of travel, focusing on the heterogeneity of individual traveling is another challenge.

In the case of ioki the models should be created on per user (traveller) basis and should take into account not only general traffic conditions and weather data but also user preferences for a type of further transportation and desired speed of the service. Knowledge on 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.

Another challenge is the integration of new on-demand transportation services (shared services) like car-sharing and bike-sharing. In particular the one-way configuration (the car or bike can be left at the destination and not necessarily at the initial pick-up point), is the one that allows for user flexibility, but represents a major challenge for the vehicle fleet optimization and relocation strategies[129].

The effect of autonomous vehicles is another aspect that needs to be considered also by ioki. Research suggests that autonomous cars can rebalance themselves in the network and coordinate their actions at a system-wide level[130], solving some of the possible system level problems of car-sharing and Litman suggests that automatic car-sharing/taxi schemes will become a reality in 2030-40s suggesting a positive impact on the on-demand services[131].

Both supply and demand models have evolved from static to dynamic capturing travel behaviour in terms of time-dependent conditions and information, and from an aggregate to a disaggregate representation of travel, focusing on the heterogeneity of individual traveling[132].

---

[127] J. G. Lee and M. Kang, "Geospatial Big Data: Challenges and Opportunities," *Big Data Res.*, vol. 2, no. 2, pp. 74–81, 2015.

[128] B. Chen and H. H Cheng, *A Review of the Applications of Agent Technology in Traffic and Transportation Systems*, vol. 11. 2010.

[129] E. Cepolina, A. Farina, and A. Pratelli, "Car-sharing relocation strategies: a state of the art," 2014, pp. 109–120.

[130] R. Zhang, K. Spieser, E. Frazzoli, and M. Pavone, "Models, algorithms, and evaluation for autonomous mobility-on-demand systems," *Proc. Am. Control Conf.*, vol. 2015–July, pp. 2573–2587, 2015.

[131] C. Electronic, "Autonomous vehicle implementation predictions : implications for transport planning Implications for Transport Planning," 2013.

[132] B. Chen and H. H Cheng, *A Review of the Applications of Agent Technology in Traffic and Transportation Systems*, vol. 11. 2010.

9.2.4 Big data dimensions and assessment

The below table provides a more detailed description of the 4 Big Data characteristics with respect to the two use cases.

Table 31 Classification according to the "4 V" of Big Data

Big Data Characteristic	Analysis for ioki
Volume	***Moderate*** – data sizes in megabytes per day. Mainly operational real traffic data from different ioki applications.
Velocity	***Moderate -*** real time operational data such as GPS location data and user booking data. It is expected to become high as more and more data and services are integrated into the ioki platform.
Variety	***Moderate*** – Structured data in the form of data descriptions and real time operational data such as route history using GPS data and user booking data. Semi-structured data as GPS coordinates, other traffic characteristics as well as event and weather data. Unstructured data in form of sensor data from vehicles and traffic monitoring devices.
Veracity	***Low*** – only the quality and trustfulness of data obtained by external sources (event and weather data) needs to be checked.

9.3 BUSINESS MODEL

There is not one single business model for MOD services, but rather, depending on the nature of the service, numerous different business models that meet the different needs of consumers, service providers and partners[133]. Basically, the models can be divided into four simplified groups:

- B2C: This is about the best possible access for private consumers to company transport services such as a fleet of vehicles, bicycles, scooters or other means of transport through memberships, subscriptions, user charges or a combination of price models, often implemented in an app.
- B2G: Provision of transport services to a public body. Pricing may include a service contract, transaction basis or other pricing model.
- B2B: Access to transport services or infrastructure for business customers either via a service charge or via user charges. The service is usually offered to employees for work-related travel.

[133] S. Shaheen, A. Cohen, B. Yelchuru, and S. Sarkhili, "Mobility on Demand Operational Concept Report."

- P2P mobility marketplace: Maintenance of a marketplace, usually an online platform between individual buyers and sellers of mobility services for a transaction fee. The platform typically provides insurance and user reviews/ratings to facilitate transactions.

A business model for MOD services like the ones offered by ioki is described below. Ioki is offering B2B services. Here for we are not only refereeing to our case study partner and their mobility service but tried to give a more universal business model for services of this kind.

In the following we describe each part of the canvas briefly:

- Value proposition. The extension of existing mobility offers via shuttle service. This offer includes fast and passenger-individual mobility.
- Customer segment. The customer segment can be distinguished between public transport providers, cities, and fleet operators.
- Customer relationship. Ioki mainly links in a digital manner with their customers. Although, there is also a direct customer interaction providing consulting personal.
- Channels. As Ioki mainly interacts on digital to deliver the value proposition to their customer segment the channels are a Smartphone application and a website.
- Key activities. The key activity is running mobility platform to offer the mobility on demand service.
- Key resources. The key resources are customer data, vehicle, and route data. These are necessary to provide their service.
- Key partners. The key partners of Ioki are public transport organisations, cities, and infrastructure owner.
- Cost structure. The major cost categories are marketing, infrastructure, IT, and personnel expenditures.
- Revenue streams. The revenue is received by service fees of the customers.
- Internalisation of costs and benefits. Implementation of data regulations, transport regulations.

<table>
<tr><th>Key partners</th><th>Key activities</th><th>Value Proposition</th><th>Customer relationships</th><th>Customer segments</th></tr>
<tr><td rowspan="3">Public transport, cities, infrastructure owner</td><td>Running mobility platform</td><td>Extension of existing mobility offers via shuttle service, fast and passenger-individual mobility offer</td><td>Digital, consulting personal</td><td rowspan="3">Public transport providers, cities, fleet operators</td></tr>
<tr><td>Key resources</td><td>Internalization of costs and benefits</td><td>Channels</td></tr>
<tr><td>Costumer data, vehicle and route data</td><td rowspan="3">Implementation of data regulations, transport regulations</td><td>App, Website, digital Interactions</td></tr>
<tr><td colspan="2">Cost structure</td><td colspan="2">Revenue streams</td></tr>
<tr><td colspan="2">Marketing, infrastructure, IT, personell</td><td colspan="2">Service Fees</td></tr>
</table>

Figure 36 Business Model Canvas for MOD service providers

9.3.1 Business processes

To better understand business processes in the context of MOD, which are various and very different, an overview of the supply and demand side is given in the following[134].

MOD supply

The supply side includes all organisations, operators, services and infrastructures involved in the provision of MOD services to persons or the supply of goods:

- Local public transport (trains, buses, etc.);
- Non-public transport services such as taxis, car rentals, journeys by private vehicles or private company fleets (Moia, Lyft, Clever Shuttle, Uber, etc.) and others;
- Goods delivery services such as logistics, first- and last-mile goods delivery, etc.
- Infrastructures such as car parks, tolls, roads and motorways;
- Vehicles of all kinds, such as private vehicles, lorries and emergency vehicles, which may be connected and/or autonomous in the future;
- Traffic management and information systems, such as parking payment systems, tolls and public transport, signalling systems, mobile apps for travel planning and payment, fleet management systems and navigation systems;
- Traffic information services including timetable information and maps such as Google maps, etc;

MOD demand

The demand side includes all users of the system (travellers and couriers), their needs and preferences, which in turn affect the supply side:

- all private and business travellers, i.e. pedestrians, passengers, drivers, cyclists; these are made up of the most diverse groups of the population, with different characteristics, including older adults, people with disabilities, children, etc;
- Goods to be delivered;
- Travel time and/or delivery conditions, which also affect the decision on the type of use;
- Origin-destination requests, determine the route and the choice of mode;
- Demand based on occupancy, size or type of vehicle requested;
- User needs and preferences, including modes and choices of how to travel (e.g. decisions to travel alone, carpooling, public transport or any other form of sharing).

Simplified, the user of the MOD service makes a route request, and this is executed by the service optimized according to the respective preferences of the user.

Three different possible mobility offers of ioki are schematically represented in Figure 37.

1. left: The mobility service connects areas or destinations that are located between the axes of the existing public transport system. Users can also be transported from home to the nearest public transport axis.

[134] S. Shaheen, A. Cohen, B. Yelchuru, and S. Sarkhili, "Mobility on Demand Operational Concept Report."

2. center: The mobility service takes users directly to larger public transport lines without the need for long detours or loops, as buses often do in peripheral urban or rural areas to include as many areas as possible in their route.
3. right: Instead of a branched bus line departing from large areas to take passengers to the next larger traffic junction, the mobility service can serve individual routes directly.

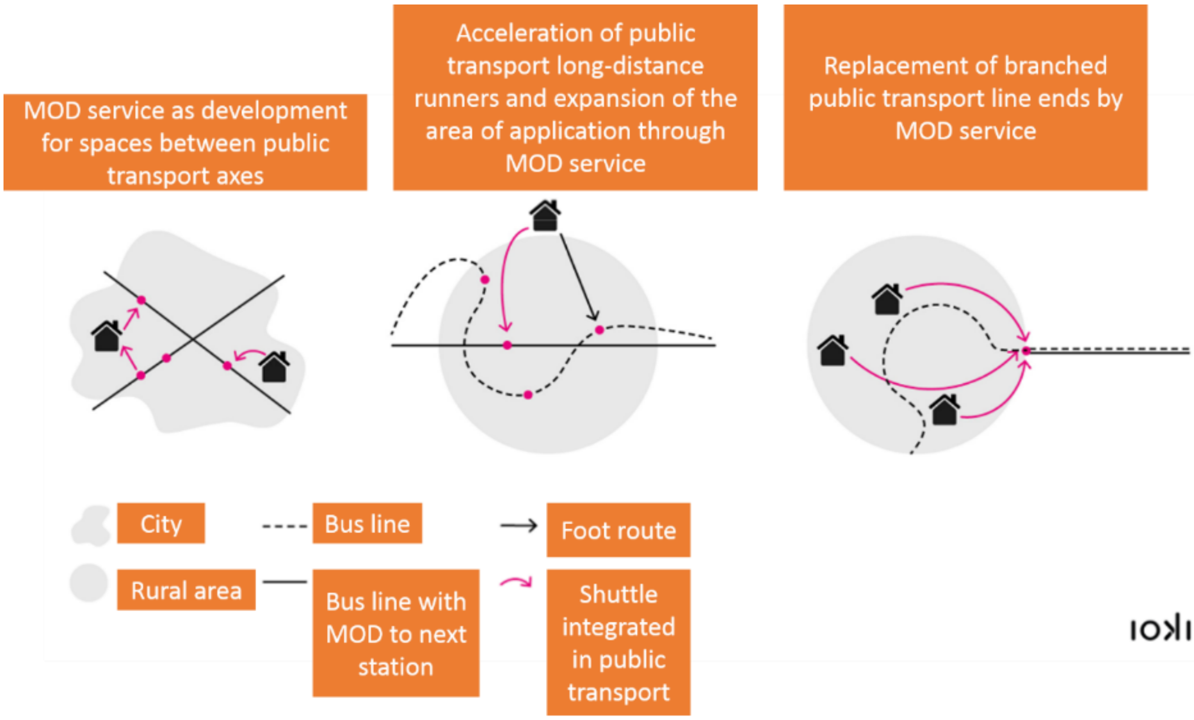

Figure 37 Schematic overview of MOD services in relation to public transport[135]

ioki's offer differs significantly from other existing on-demand services. Usually, new on-demand services are offered in city centres, where they compete directly with public transport or foot races. This is understandable, as customers are available there at almost any time of the day. However, these services counteract current political goals such as car-free, less noisy and emission-free inner cities. This way the existing public transport network is not being strengthened but is coming under pressure and losing customers.

ioki does not offer any traffic in inner cities, but rather connects rural areas with poor connections. The aim is to enable a person to travel the first and last mile integrated into the existing local public transport network. It works according to an asset-light model, i.e. they develop the service and its feasibility, but does not produce or own vehicles or other infrastructure themselves. Customers wishing to implement ioki's offer need, among other things, appropriate vehicles (5-6-seater buses with space for prams), trained drivers and, in the case of an electric vehicle fleet, the appropriate charging infrastructure.

[135] D. N. Koblenz, “ioki presentation.”

In addition to already mentioned mobility services, ioki works on the implementation of autonomous driving services. They are currently being developed separately and will be merged when development has progressed far enough.

9.3.2 Relation to big data initiatives

Ioki enriches their available data with different open data initiatives examples are:

- **Open government data** to achieve socio-economics input and extend their market analysis.
- **Open street map** data to enrich their knowledge about the existing infrastructure they can use.
- Other public available **open data** like weather data to extend their models with more relevant information.

9.3.3 Illustrative user stories

Example 1, MOD service experience[136]:

"Last week I had a meeting in the centre of Amsterdam. When I arrived at train station Zuid, Google showed me that there were massive traffic jams in the city centre (like always) so Uber was not an option. The tram wouldn't stop anywhere near my destination, so I decided to rent a bike. Within 10 minutes I reached my destination. If I had gone there by car it would have taken me a lot longer to get there and it would have been more expensive as well!"

Example 2, ioki in Hamburg:

In Hamburg, ioki services are offered in the districts of *Lurup* and are offered in cooperation with the local transport company Verkehrsbetriebe Hamburg-Holstein GmbH (VHH). Users here are almost exclusively public transport customers who also pay with their tickets. This can be selected as a payment method in the associated ioki app and is well used. These districts are poorly connected by the existing public transport system. However, they are densely populated residential areas with a great need for mobility, and there are few good alternatives to public transport. The shuttle service, for example, takes users to the next larger public transport station where they again have a good connection. So, journeys go from an address given by the customer to a public transport station and vice versa. In Hamburg, customers are transported with electric 5-6 seated vehicles.

Example 3, route of the autonomous driving shuttle of ioki:

For the development of an autonomous driving service, Ioki has obtained the first registration of an autonomous driving vehicle in Germany. This runs on a 1.2 km long test track in Bad

[136] https://smartstories.nl

Birnbach[137]. There it connects a central bus station with the local spa and is used by passengers, which are mainly the customers of the spa. There is still always an operator in the vehicle who can intervene if necessary. The vehicle operates on a permanently programmed route, every small deviation from this programmed line stops the vehicle and requires the manual intervention of the operator. At this point the autonomous shuttle is designed for maximal safety, but not very flexible.

9.4 ANALYSIS OF SOCIETAL ASPECTS

9.4.1 Economic and political

In the following a list of possible economic and political issues due the implementation of MOD services is given:

- Traditional on-demand mobility offers, like the conventional taxi services or call-bus systems might be shrunken or replaced.
- Existing regulations concerning driver trainings, vehicle standards or pricing regulations need to be updated.
- Rural areas might get even more unattractive to live in, if mobility is easily and cheap available in cities and people are not willing dispense on this kind of life quality.
- In urban areas MOD services might support the following mobility needs: goods delivery in first and last mile, first and last mile transport for passenger to public transport, mobility services late at night or other hours of low offer.

According to our interview partners, most European rail operators have similar departments like Smart City. They do not compete as rail transport is still a national business in many respects. However, there is a close communication with railway-partners, like the Swiss, Austrian or French railway.

Smart City cooperates with German cities such as Hamburg. There is a cooperation agreement that stipulates that the concepts developed by Smart City over the next four years will be strategically implemented in Hamburg in partnership with the city. There are also further contacts to many different cities (various in size) in Germany.

137 "Testroute autonomous driving." [Online]. Available: https://www.deutschebahn.com/en/Digitalization/auto_driving/testfield_birnbach-1215180. [Accessed: 05-Apr-2019]

9.4.2 Social and ethical

There are the following five main challenges concerning social and ethical aspects of MOD[138][139]. The mentioned studies were deduced in the US-American market but may apply to the European situation as well.

1. **Discrimination against certain groups in society**. Access and equal treatment of minorities, women and people with disabilities must be ensured for MOD services. For example a study conducted in 2016 in several US cities showed racist behaviour among Lyft and Uber drivers[140].
2. **Access and usability for older adults and people with disabilities.** Another challenge is to adapt MOD services to the needs of older people and people with disabilities. ioki, for example, operates its Hamburg fleet with vehicles that offer sufficient space for wheelchairs and a ramp that allows access to the vehicle[141].
3. **Digital divide.** While some MOD services can be accessed without a smartphone, many on-demand travels planning, booking and payment services generally require a mobile device connected to the Internet. Lack of mobile Internet access can prevent users from participating in many MOD services.
4. **Connection in rural areas**. Many new MOD services start in urban areas with secured demand. The availability of these services in rural areas, at least in part, is a major challenge. Often, existing local transport solutions such as buses are no longer fully utilised in rural areas, so that there is a definite need for new and intelligent mobility solutions. Especially in such areas autonomous on-demand services can help. ioki is implementing exactly this in an edge-city area of Hamburg, as it better connects the area with the public transport network.
5. **Economical accessibility.** A challenge for MOD services is not to deny access to individual groups since payments are often made digitally.

The abovementioned challenges mainly relate to data-driven social discrimination. This issue was also identified as one of seven key ethical and social issues in Deliverable D2.3.

There is namely fear of the loss of jobs, especially among drivers in public transport, but also of delivery traffic, triggered by new on-demand services, especially autonomous driving services.

On the other hand, many jobs are currently unfilled and there are too few drivers, especially in inner-city delivery traffic. The current drivers are often under a lot of pressure because they have a very high workload and deliveries are not centralized.

According to our interview partners, social and ethical issues are considered by DB's Strategy Department and are no longer given special, standardised attention in the individual departments. The topic of "digital divide", i.e. possible social problems arising from lack of

---

[138] S. Shaheen, A. Cohen, B. Yelchuru, and S. Sarkhili, “Mobility on Demand Operational Concept Report.”

[139] A. Rosenblat, K. E. C. Levy, S. Barocas, and T. Hwang, “Discriminating Tastes: Uber’s Customer Ratings as Vehicles for Workplace Discrimination,” *Policy and Internet*, vol. 9, no. 3, pp. 256–279, 2017.

[140] S. O’Brien, “Black riders wait longer for Uber rides, study reveals,” 2016.

[141] Ioki and Vhh, “ioki Hamburg.” [Online]. Available: https://vhhbus.de/ioki-hamburg/. [Accessed: 17-May-2019].

access to digital media or lack of competence, is also not considered separately in the concepts and business models.

Additionally an analysis of the societal and ethical issues is done according to the six topics identified by Deliverable 2.3 of the LeMO Project.

Table 32 Societal and ethical issues in MOD services

Aspect	Description of aspect	Relevance of issue to Ioki
Trust	Increase of decrease in trust in by or between the data users, data facilitators and data suppliers.	No issues related to trust were found.
Surveillance	The ability to collect large amounts of data enables the data user to better understand processes in society.	Ethical implications may arise from the linking of personal data with process data.
Privacy	The involvement of privacy sensitive information	The collection of processing data in combination with the personal user information could allow detailed information about movement profiles, peer groups, working places, etc.
Free will	Big data analysis and associated predictive analysis that threatens human autonomy	Suggestions or personalized prizing by the MOD could influence the decision making of the users to take a certain connection and / or transport vehicle.
Personal data ownership	Confusion of who owns the data or loss of data ownership by data suppliers	The personal user data should only be operated at the MOD provider and within the necessary processes.
Discrimination	Discrimination of population groups based on generalisations derived from big data	Access and equal treatment of minorities, women and people with disabilities must be ensured for MOD services.

9.4.3 Legal

Ioki is heavily dependent on open data, such as maps and weather data, and real-time operational data, such as GPS location data and user booking data. The collect data from the following sources: (i) geo-data, (ii) the DB traffic model and (iii) the Open Data Portal of DB.

These data are collected using a geo-information system, software for traffic models as well as self-developed algorithms.

As real-time operational data can also include personal data, notably of customers of Ioki, it has dealt with legal requirements on privacy and personal data protection and has implemented relevant security measures.

To analyse the legal issues, the framework from D2.2 is used and can be found in the below table. This analysis explicitly refers to our interview partner ioki, since the diversity of MOD services does not allow any concerted statement on the individual aspects.

Table 33 Legal aspects analysed for ioki

Aspect	Description of aspect	Relevance of aspect to ioki
Privacy and data protection	Consideration and implementation of the requirements of the GDPR and other types of privacy and data protection-related legislation	With the introduction of the General Data Protection Regulation (EU) 2016/679, non-compliant tools were replaced by new ones. A corresponding role and authorization system for personal data was introduced.
(Cyber-)Security	Awareness and implementation of certain general or specific (cyber-)security requirements	Ioki is aiming for certification according to the ISO 27001 standard [6]. This meets very high standards of data security and protection and is perceived as an additional feature for quality of service.
Anonymization and pseudo-nymisation	The use of any type of anonymisation or pseudonymisation technique, i.e. techniques to ensure personal data can no longer be attributed to individuals without using additional information or to reduce the identifiability of individuals	Data, which is used for general analysis like market research, is anonymised and pseudonymised. A clear hierarchy of access rights to personal customer data in accordance with data protection regulations has been implemented.
Open data	Awareness and use of open public sector information	Open data is used for market analysis. Examples are socio-economic data of open government portals or publicly available maps.
Data sharing obligations	Awareness of and compliance with certain specific data sharing obligations, mandatory legal requirement to share data with other market actors	Data sharing is often defined via contractual agreements.
Data ownership	Third parties claiming data ownership	The ownership of collected user and vehicle data is defined in contracts between ioki and their partners.

The German Passenger Transport Act[142] is one of the most comprehensive in Europe and offers some challenges for the implementation of modern on-demand transport. This law regulates the commercial transportation of people in e.g. taxis, buses, trams, etc.

For example, so-called positioning journeys are legally restricted, i.e. event data may not be used without further ado. Now, special permits or exemptions are granted for on-demand services.

As an example, an on-demand mobility provider operates currently in Hamburg with an "experiment clause", which was granted for 5 years. This will then be withdrawn, and the service or regulation may have to be adapted. Steps to improve this law and adjust it to modern developments are seen as necessary soon.

The implementation of this nationwide law is also left to the municipalities to interpret, i.e. uniform implementation is not guaranteed. In Hamburg, an on-demand mobility provider waited 6-8 months for a permit to operate a vehicle, in Berlin they were completely rejected. Partly the traceability is not given.

9.4.4 Environmental

According to a study, which analyzed the user behavior of free-floating car sharing users in several German cities under environmental aspects, this relatively new service does not have a significant positive effect on the climate and greenhouse gas emissions[143].

Free-floating car sharing is an example of a relatively new MOD service. The carsharing vehicle can be booked digitally, via Internet or App or by telephone. While in stationary car sharing the cars have fixed parking spaces where they have to be picked up and parked again after use, in flexible concepts (free-floating car sharing) the vehicles are available for spontaneous use in public spaces.

The use of free-floating car sharing had little influence on decreasing CO2 emissions. The analyses showed that compared to the time before registration for car sharing, not less but more distances were covered by car. However, in one city, where the car sharing fleet consists of electric vehicles, a reduction of emissions could be shown. The use of car sharing had no effect on the use of local public transport. In addition, only a few of the respondents abolished their private car in favor of car sharing. Only about three percent abolished their car, so that the study results do not point to any significant contribution to relieving public road space through free-floating car sharing.

To summarize some important aspects that a MOD service must consider, to have a positive climate impact:

---

[142] "German Passenger Transport Act." [Online]. Available: https://www.gesetze-im-internet.de/pbefg/BJNR002410961.html. [Accessed: 05-Apr-2019].

[143] A. Rosenblat, K. E. C. Levy, S. Barocas, and T. Hwang, "Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination," *Policy and Internet*, vol. 9, no. 3, pp. 256–279, 2017.

- The total distances covered by passengers or goods should not increase due to a MOD service (e.g. avoiding rebound effects[144]);
- The MOD service should be carried out using modes of transportation, which rely on renewable energy to reduce total emissions;
- The MOD service should in the best case replace vehicles that are privately owned and reduce the individual car ownership rate;
- The MOD service should not replace existing public transport offers, that are already more environmentally friendly, like trains, buses, etc.

9.5 ASPECTS RELATED TO OPPORTUNITIES, BARRIERS AND LIMITATIONS

9.5.1 Opportunities

The main opportunities, which were identified based on the expert interviews and the related research, are listed below:

- Connect ex-urban, rural or not well-connected areas with existing public transport system;
- Decrease barriers to public transport access via offering first and last mile mobility services;
- Simplify payment procedures by offering integrated ticketing systems for different MOD service providers;
- Improving urban living conditions, like pollution, noise, space, by reducing number of privately-owned vehicles and the urban traffic;
- Reducing greenhouse emissions by implementing more efficient and environmentally friendly mobility offers like electrically operated car sharing fleets or extended public transport;
- Improving first and last mile goods and packages delivery in urban areas with intelligent and centralized MOD services, like using cargo-bikes, centralized distribution-hubs, etc.

9.5.2 Barriers and limitations

The main challenges and barriers, which were identified based on the expert interviews and the related research, are listed below:

- Some legislation is not adapted to modern on-demand transport and is therefore a barrier. Furthermore, the implementation of some nationwide laws is left to the municipalities to interpret, i.e. uniform implementation is not guaranteed. An example is the German Passenger Transport Act, which is one of the most regulative in Europe and offers some challenges for the implementation of modern on-demand transport.
- Avoiding exclusion of certain societal groups from having access to MOD services, like elderly people or people with no or less digital knowledge;

[144] H. J. Walnum, C. Aall, and S. Løkke, "Can rebound effects explain why sustainable mobility has not been achieved?," *Sustain.*, vol. 6, no. 12, pp. 9510–9537, 2014.

- Resources are needed for harmonizing different IT-infrastructures to provide integrated services, like the interoperability of different systems and service providers or to ensure the availability of e-payment and e-tickets;
- Ensuring that conventional taxi services are not replaced by new and lesser regulated services with poorly educated drivers, but create new services that support existing public transport systems and help to reach societal goals like climate protection and living-worthy urban areas.

9.6 SOURCES OF INFORMATION

9.6.1 Interview

Interviews were carried out with the following:

- The head of Digital Innovation & Strategy at Smart City DB
- Head of Projects & Operations Management at Ioki

9.6.2 References

ioki, "ioki - Inspiring smart mobility," 2019. [Online]. Available: https://ioki.com/. [Accessed: 05-Apr-2019].

D. N. Koblenz, "ioki presentation."

S. Shaheen, A. Cohen, B. Yelchuru, and S. Sarkhili, "Mobility on Demand Operational Concept Report." .

"Testroute autonomous driving." [Online]. Available: https://www.deutschebahn.com/en/Digitalization/auto_driving/testfield_birnbach-1215180. [Accessed: 05-Apr-2019].

S. Dabiri and K. Heaslip, "Transport-domain applications of widely used data sources in the smart transportation: A survey," pp. 1–53, 2018.

L. A. Klein, "Traffic Detector Handbook:Third Edition—Volume II," US Department of transportation, 2006.

"Open Data Portal DB." [Online]. Available: https://data.deutschebahn.com/. [Accessed: 05-Apr-2019].

R. Khan, S. U. Khan, R. Zaheer, and S. Khan, "Future Internet: The Internet of Things Architecture, Possible Applications and Key Challenges," in 2012 10th International Conference on Frontiers of Information Technology, 2012, pp. 257–260.

R. Petrolo, V. Loscrì, and N. Mitton, "Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms," Trans. Emerg. Telecommun. Technol., vol. 28, no. 1, p. e2931, Jan. 2017.

P. Masek et al., "A Harmonized Perspective on Transportation Management in Smart Cities: The Novel IoT-Driven Environment for Road Traffic Modeling," in Sensors, 2016.

B. Chen and H. H Cheng, A Review of the Applications of Agent Technology in Traffic and Transportation Systems, vol. 11. 2010.

N. E. El Faouzi, H. Leung, and A. Kurian, "Data fusion in intelligent transportation systems: Progress and challenges - A survey," Inf. Fusion, vol. 12, no. 1, pp. 4–10, 2011.

J. G. Lee and M. Kang, “Geospatial Big Data: Challenges and Opportunities,” Big Data Res., vol. 2, no. 2, pp. 74–81, 2015.

E. Cepolina, A. Farina, and A. Pratelli, “Car-sharing relocation strategies: a state of the art,” 2014, pp. 109–120.

R. Zhang, K. Spieser, E. Frazzoli, and M. Pavone, “Models, algorithms, and evaluation for autonomous mobility-on-demand systems,” Proc. Am. Control Conf., vol. 2015–July, pp. 2573–2587, 2015.

C. Electronic, “Autonomous vehicle implementation predictions : implications for transport planning Implications for Transport Planning,” 2013.

K. Hee et al., “Big Data Methodologies, Tools and Infrastructures,” 2018.

R. Macário, “TURBLOG: Conceitos e Modelos de Negócio para a Logística Urbana,” 2011.

A. Rosenblat, K. E. C. Levy, S. Barocas, and T. Hwang, “Discriminating Tastes: Uber’s Customer Ratings as Vehicles for Workplace Discrimination,” Policy and Internet, vol. 9, no. 3, pp. 256–279, 2017.

S. O’Brien, “Black riders wait longer for Uber rides, study reveals,” 2016.

Ioki and Vhh, “ioki Hamburg.” [Online]. Available: https://vhhbus.de/ioki-hamburg/. [Accessed: 17-May-2019].

J. Debussche, J. César, M. Hong, Z. Nordeng, and R. Waldenfels, “Report on ethical and social issues,” 2018.

J. Debussche, J. César, and I. De Moortel, “Report on legal issues,” 2018.

“German Passenger Transport Act.” [Online]. Available: https://www.gesetze-im-internet.de/pbefg/BJNR002410961.html. [Accessed: 05-Apr-2019].

H. J. Walnum, C. Aall, and S. Løkke, “Can rebound effects explain why sustainable mobility has not been achieved?,” Sustain., vol. 6, no. 12, pp. 9510–9537, 2014.

J. Debussche, J. César, M. Hong, Z. Nordeng, and R. Waldenfels, “Report on ethical and social issues,” 2018.

10 Conclusions

The case studies carried out within Task 3.2 cover a diverse selection of innovative applications of BDT in the transport sector. Using a consistent but flexible methodology, these case studies describe the complexity of using big data in the transport sector.

Case study 1 focused on the application of big data technologies in predictive maintenance in rail transport. The digitalization of train and rail infrastructure operations produces valuable data, which when analysed with the right big data technologies, can improve the prediction of potential train-related failures and the precision of maintenance schedules, thus improving overall safety and fleet availability. As rail transport forms the backbone of the transport strategy of many Member States in the EU, any promising improvement in railway operations will have a considerable positive impact on the EU as a whole.

Case study 2 explored the opportunities and challenges of successfully making big data available as open data. Open data is an important policy objective and strategy of the European Commission to support research and innovation in the field of transport. While it may seem simple to implement, making big data open entails many technological and organisational challenges, especially when economic incentives are lacking. Initiatives on open data are currently championed by transport authorities and organisations that rely on data, especially to keep up with customer demands and to provide innovative transport services through real-time and precise transport information.

Case study 3 explored the real-time traffic management which is one of important parts in Intelligent Transport Systems (ITS) that rely on big data analytics. ITS and the emerging field of C-ITS (Cooperative Intelligent Transport Systems) is effective in dealing with many challenges of managing road traffic, as well as in dealing with the emergence of connected and automated vehicles. While many transport management systems in Europe produce a wide variety and volume of data, there is still an untapped potential for the widespread use of big data techniques to improve traffic flows.

Case study 4 provides an in-depth description and analysis of the use of big data to support logistics processes in an era where customer and supply chain requirements are stricter and more varied. The study is based on the very advanced and successful company Kepler51, which is based in the USA. The case study presents their big data solution that monitors, assesses, and forecasts transport risks and delays along the transport chain. However, their application, which focuses on non-vehicle data as a primary source, does serve to illustrate some of the potential benefits and risks facing transport carriers in Europe when considering the use of big data technology.

Case study 5 introduces a unique application of big data to coordinate the traffic of both the road and inland waterway network in the Netherlands. Coordination of inland waterways is an important topic in the Netherlands, especially since the network intersects with the road network at bridges. The major challenge faced was the political challenge to organise the collaboration between powerful governmental bodies and port authorities. But, once the initial challenge was met, the benefits of the programme were acknowledged, spurring replication in other parts of the Netherlands. While other countries in the EU may not have a similar application (i.e. the integration of the road and inland waterway network), the lessons learned in this case provide important clues as to the organisational challenges of multi-stakeholder large-scale project implementation.

Case study 6 presents the application of big data technology and analysis methods to provide a multimodal public transport information and route planning service to passengers. The challenge has been to develop a service that provides real-time information to passengers, while preserving the brand and identity of the individual public transport operators. This also offers an important opportunity for identifying an alternative to current non-European service providers that reduce the digital autonomy in Europe, thus leading to an imbalanced market. The service discussed in the case-study supports improved mobility and customer experience by providing effective route planning by combining open data from various sources. It also intends to further improve its product in the future by combining data collected from end-users (e.g. through data from fare cards), but this seems difficult on account of privacy concerns expressed by end-users. The case serves as an important source of knowledge about the issues that many cities are facing in terms of both improving the multimodal transport information services as well as route planning, especially in the context of disruptive and innovative transport services (e.g. scooter sharing).

Case study 7 presents the challenging and disruptive on-demand urban mobility services sector. A disruptive technology company working in the heavily regulated urban transport sector needs to be innovative in order to tackle the challenges of policy, a small customer base, and the competition with established firms. Furthermore, the technologies supporting the activity have to be very reliable and attractive so that new customers easily become regular customers. The case presents a spin-off from an established company aiming to cover the gaps in their existing services. Big data, especially open data, empowers the start-up to provide niche services that fill the gaps in the existing public transport system.

The case studies provide a realistic view of the opportunities and benefits provided by big data in the specific domains. However, they also bring to light the very many challenges that must be overcome for a successful and sustainable improvement of transport sector using big data. The issues highlighted must therefore be carefully addressed via policy at the right governmental levels and via further research to improve the technical application of the technology. The LeMO consortium will rely on the results of these case studies in the next phase of the project in Work Package 4 to conduct the horizontal analysis and road-mapping exercises.

In particular, the cases have explored in-depth current issues that may pose as limitations and barriers for replication of big data-based innovations in business models. These proceed from the technological and technical limitations, and that which are affected by economic, political issues, legal, ethical and social issues currently facing the market. Some of the case studies have also dealt with some of these issues successfully and offer guidance on how to address other potential issues that may arise. These limitations, opportunities and barriers will be listed and elucidated in Work Package 4, paving the way for the creation of a research and policy roadmap for the European Transport. The roadmapping exercise will be conducted jointly with key stakeholders and experts in the field.