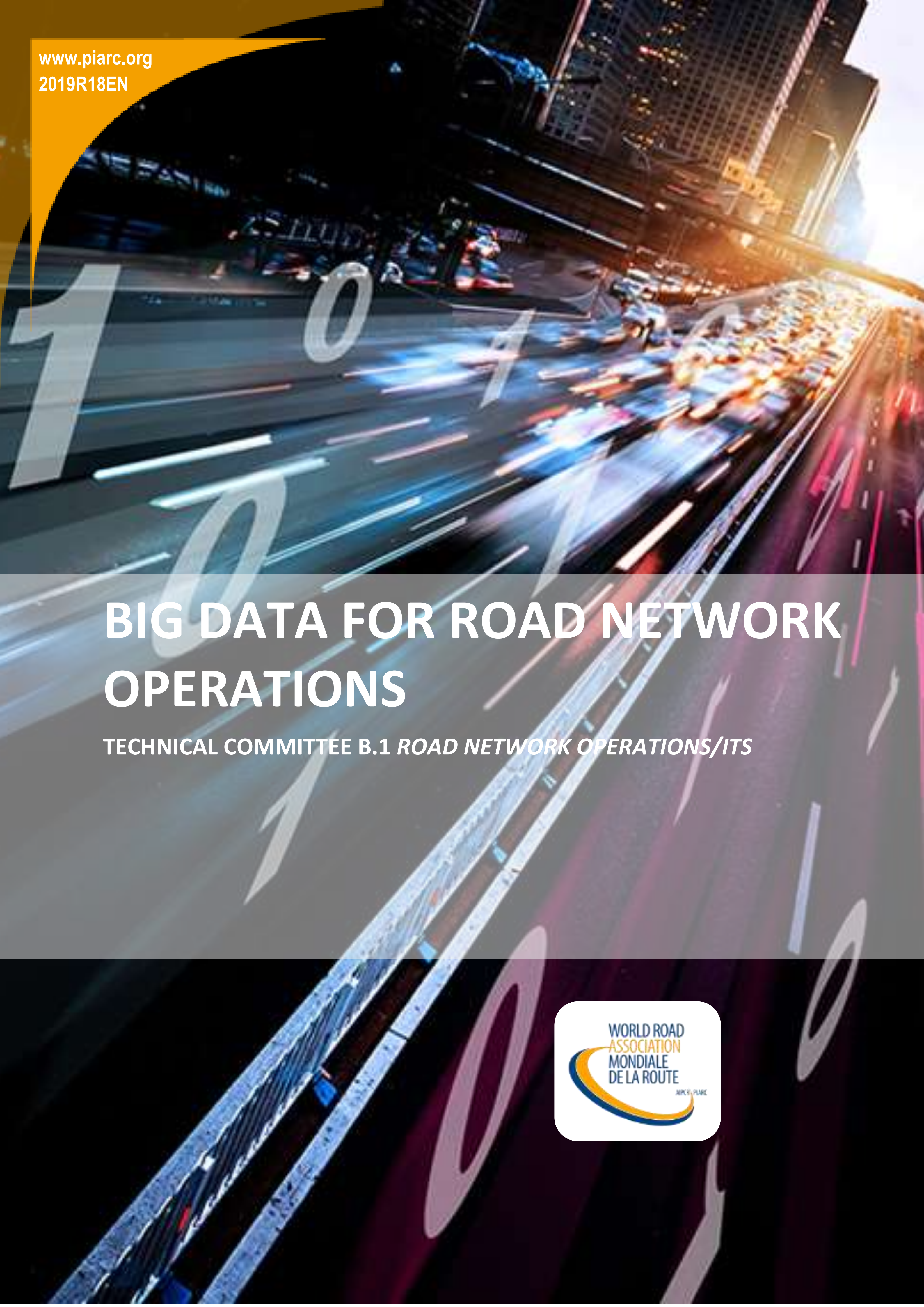
www.piarc.org
2019R18EN



BIG DATA FOR ROAD NETWORK OPERATIONS

TECHNICAL COMMITTEE B.1 *ROAD NETWORK OPERATIONS/ITS*

WORLD ROAD ASSOCIATION
MONDIALE DE LA ROUTE
AIPCR · PIARC

STATEMENTS

The World Road Association (PIARC) is a nonprofit organisation established in 1909 to improve international co-operation and to foster progress in the field of roads and road transport.

The study that is the subject of this report was defined in the PIARC Strategic Plan 2016– 2019 and approved by the Council of the World Road Association, whose members are representatives of the member national governments. The members of the Technical Committee responsible for this report were nominated by the member national governments for their special competences.

Any opinions, findings, conclusions and recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of their parent organisations or agencies.

This report is available from the internet site of the World Road Association (PIARC): http://www.piarc.org

World Road Association (PIARC)
Arche Sud 5° niveau
92055 La Défense CEDEX, FRANCE

International Standard Book Number: 978-2-84060-559-1

Front cover © Technical Committee B.1 Road Network Operations/ITS

BIG DATA FOR ROAD NETWORK OPERATIONS

TECHNICAL COMMITTEE B.1 *ROAD NETWORK OPERATIONS/ITS*

AUTHORS/ ACKNOWLEDGEMENTS

This report was prepared by the working group *B.1.3* of the Technical Committee B.1 Road Network Operations/Intelligent Transport Systems of the World Road Association (PIARC).

The contributors to the preparation of this report were:

- Andreas Rau (Germany)
- Arve Kirkevold (Norway)
- Daniel Allaire (Canada/Quebec)
- Dieter Hintenaus (Austria)
- Emmanuelle Freneat-Laleu (France)
- Ian Espada (Australia)
- Keechoo Choi (South Korea)
- Lena Radics (Austria)
- Marek Scerba (Czech Republic)
- Martin Böhm (Austria)
- Yongju Yi (South Korea)

The editors of this report were Martin Böhm (Austria) and Ian Espada (Australia) for the English version.

The translation into French of the original version was produced by France. Daniel Russomano (Argentina) translated the original version into Spanish.

Martin Böhm (Austria) was responsible within the Technical Committee of the quality control for the production of this report.

The Technical Committee was chaired by Jacques Ehrlich (France). Sylvain Belloche (France), Valentina Galasso (Italy), Daniel Russomano (Argentina) were the French, English and Spanish-speaking secretaries, respectively.

EXECUTIVE SUMMARY

2019R18EN

BIG DATA FOR ROAD NETWORK OPERATIONS

Currently, several technological developments are influencing the day-to-day work of road network operators. Automation, internet of things (IoT), smart systems, mobility services, artificial intelligence and big data are aiming to improve traffic operations and enhance mobility. Among these technologies, big data and big data processing technologies are key enablers for all other technologies. Big data has the potential to influence the whole value chain in the road transport sector.

The scope of this report is the collection of big data and the description of the framework for big data management and usage based on global best practice. In particular, the focus is on using big data for internal processes related to road network operations. The report begins with a description of the definition of big data in the transport domain. There is no globally accepted definition of big data, but there are three commonly identified characteristics of big data that are particularly valid for the transport domain, i.e. large volume of data; high velocity of data collection; and, variety of data which consists of different formats usually from varied sources. The big data process from data collection to the generation of services and associated knowledge is then described. The potential of big data in transport is also identified and analysed.

Chapter three focuses on the future potential of big data for road network operations, with particular focus on integrated services that are enabled by big data. Services such as Cooperative, Connected, and Automated Mobility (CCAM) and Mobility as a Service (MaaS) are discussed. Risks in big data in the context of processing and applications are also examined.

Chapter four elaborates on the data. There are different types of data. Data also have different timelines, i.e. static, real-time, dynamic and historical. Data also comes from a variety of sources (including detectors, sensors, third party, or social media data). Given the variety of data and varied sources of data, a well-developed metadata plays an important role to maximise the usability of the data.

Chapter five presents use cases in the highway and public transport domain, with particular attention to internal use of big data processing for road network operation. The discussion covers data access infrastructure; traffic management; road network operation services; automated weather detection services; and, planning and operations. Possible future applications in the areas of road planning, design and operations are also discussed. Applications of big data in public transport operations and shared mobility are also illustrated. The chapter concludes with a review of the benefits and limitations of big data in road network operation.

The key items that need to be addressed by transport organisations to be ready for big data processing and analytics is covered in chapter six. The skills needed and standards to be followed are analysed. This chapter also illustrates a big data reference architecture, which focusses on organisational requirements. A discussion on quality, data ownership, security, and privacy concludes the chapter.

Chapter seven summarises the findings of the report. It also includes recommendations on how to deal with big data in the road network operation domain.

CONTENTS

1. OVERVIEW OF THE INPUTS GATHERED 4
2. INTRODUCTION 5
2.1. WHAT IS BIG DATA IN DRANSPORT 6
2.2. FROM DATA TO SERVICES 7
2.3. THE BIG DATA CYCLE 8
3. BIG DATA ENABLING FUTURE MOBILITY 11
3.1. RELATION OF BIG DATA WITH OTHER TECHNOLOGY TRENDS 11
3.2. INTEGRATED SERVICES ENABLED BY BIG DATA 12
3.3. THREATS 16
4. IT IS ALL ABOUT DATA 17
4.1. WHAT IS DATA 17
4.2. DATA EVOLUTION 17
4.3. DATA AND TIMELINE 17
4.4. DATA SOURCES 20
4.5. METADATA 25
4.6. MOBILITY DATA MARKET PLACES ENABLING BIG DATA ACCESS 26
5. APPLICATIONS 28
5.1. ROAD CASE STUDIES ON PROVEN AND EXISTING APPLICATIONS 28
5.2. ROAD CASE STUDIES ON EXPECTED FUTURE APPLICATIONS IN THE AREAS OF PLANNING; DESIGN; AND OPERATIONS 34
5.3. PUBLIC TRANSPORT CASE STUDIES 36
5.4. SHARING MOBILITY 41
5.5. BENEFITS IN PLANNING, DESIGN AND OPERATION FOR ROAD NETWORK OPERATION 43
6. ENABLING ELEMENTS 48
6.1. MAKING ORGANISATIONS READY (SKILLS NEEDED) 48
6.2. STANDARDISATION 52
6.3. PROCESS MODEL (BIG DATA REFERENCE ARCHITECTURE) 52
6.4. QUALITY CRITERIA 57

6.5. DATA OWNERSHIP 58
6.6. PRIVACY 58
6.7. SECURITY 61
7. CONCLUSIONS 63
8. KEY RECOMMENDATIONS 65
9. GLOSSARY 66
10. REFERENCES 69
APPENDICES 71
10.1. ANNEX A: OVERVIEW OF METADATA ELEMENTS FOR A METADATA SET [5] .. 71
10.2. ANNEX B: CASE STUDY CZECH REPUBLIC ON BIG DATA PROCESSING 74
10.3. ANNEX C: CASE STUDY ON AUTOMATED TRAFFIC SIGNAL PERFORMANCE MEASURES IN THE STATE OF UTAH 76
10.4. ANNEX D: CASE STUDY NETHERLANDS ON REDUCING RUSH HOUR CONGESTION 78
10.5. ANNEX E: CASE STUDY AUSTRALIA (SYDNEY) ON POST-IMPLEMENTATION ANALYSIS 80
10.6. ANNEX F: CASE STUDY LONDON ORBITAL (UK) ON THE USE OF NEAR REAL TIME DASHBOARDS TO IMPROVE OPERATIONS PERFORMANCE ON THE M25 LONDON ORBITAL MOTORWAY 82
10.7. ANNEX G: CASE STUDY JAPAN ON ELECTRONIC TOLL COLLECTION 84
10.8. ANNEX H: CASE STUDIES SOUTH KOREA ON PUBLIC TRANSPORT NETWORK PLANNING 86

1. OVERVIEW OF THE INPUTS GATHERED

A wide range of inputs was collected for this report. In total 44 policy documents on big data in Road Network Operations were collected based on desk research and personal knowledge of the working group experts.

An in-depth analysis of these reports formed the baseline for the preparation of the Technical Report. The analysis was followed by seven intensive workshops with the working group experts that brought in their country- and region-based knowledge in bi-annual meetings of the Technical Committee B.1. In addition, eight web meetings were conducted to discuss content and progress in between the face-to-face workshops.

The policy reference documents that have been collected and reviewed have been published in following countries:

- Australia
- Austria
- Canada
- China
- Czech Republic
- Egypt
- France
- Greece
- India
- Norway
- South Korea
- Sweden
- Switzerland
- United Kingdom
- United States of America

In addition to the Technical Report, ten case studies on big data were prepared by the technical committee. The case studies are available as Appendix to this report.

2. INTRODUCTION

Digital transformation is in full swing in all areas throughout the world and the transportation sector is not an exception. Becoming and/or transforming to 'digital' simply means 'paperless'; and, one quick example in the transport sector is the transition from paper tickets to smartcards. Another aspect of becoming digital is the inevitable accumulation of digital transport information, which may be in a smartphone, computer or cloud; and, the accumulation of significant amounts of data leads to a big data environment of transport. Everyday people use navigation, toll tags and smartphones for public transportation and taxi. Bus and freight trajectories are being monitored and accumulated on a real-time basis. This is the so-called 'transport big data environment'.

The transport big data environment have become a reality due to achievements in computing technology and the rise of consumer demands which has led to advanced applications aimed at facilitating people's lives. Digitalisation with corresponding increase in the amount of data is enabled by data storage, as well as, data processing technologies (e.g. data science, artificial intelligence, deep learning, and machine learning). New benefits arise from looking deep into the data and getting new insights and visibility on movement behaviour of people, goods and vehicles. Previously unimaginable insights in transport have become feasible.

With better insight and visibility, major aspects of the transportation sector such as planning, design and operation can be improved. Congestion, accident, and environmental degradation levels can be mitigated and new approaches to travel forecasting can be developed and tested. Owners of transport data are exploring business applications and services, and the public sector is starting to apply the data to enhance transportation planning and operations.

Data by its nature involves tools such as statistics, advanced data science and artificial intelligence. The current approaches in travel demand forecasting, design schemes, accident reduction and ITS operation will be met with a new set of challenges. New technologies in transportation are being enabled in parallel with the use of big data and associated tools, which will follow various phases of public acceptance before they are fully applied in operation.

Big data is sometimes considered under its various technical aspects such as data science, data lakes or deep learning. While certain aspects of big data are still emerging, big data has a well-understood potential to improve the transportation sector that it may be confidently concluded that big data's potential is not overhyped and is already being applied in the transport sector demonstrating its value. Such being the case, this report would examine big data including the definition of big data in the context of road network operations; possible benefits of its applications; enabling elements including standards; and, plausible solutions in addressing issues such as ownership, privacy and liability.

Infrastructure operators are keen on big data, as it promises to impact the whole value chain of the road transport sector. In this context, the following forms the fundamental hypothesis for this report:

"Big data is a ground-breaking tool for infrastructure operators, which can be applied to improve services to achieve policy goals and meet road users' expectations."

The scope of this report is the collection of big data and the description of the framework for big data management and usage based on global best practice. In particular, the focus is on using big data for internal processes related to road network operations. Best practices of key end user services based on big data are described in this report. The analysis focuses do not cover all potential applications, in particular, this report does not cover potential upcoming end-user services, such as, cooperative ITS (C-ITS), automation, and Mobility as a Service (MaaS).

2.1. WHAT IS BIG DATA IN DRANSPORT

Currently there is no globally accepted definition of big data, particularly in the context of the Transport Sector. A commonly agreed definition of "big data in Transport" and especially "big data in Road Transport" forms the basis for this report.

In general, big data is characterised by the 5 V's:

- Volume – large quantities of data

Large volumes of data in the transport sector is achieved either through spatial and/or temporal coverage (e.g. data collection by numerous sensors and/or through a long periods of time). Moreover, large quantities of data is achieved by detailed and comprehensive coverage, wherein data is collected from a sigfnicant proportion of the data population rather than from a small sample subset only. To collect these large quantities of data, big data is sourced from permanently updated sources including 3rd party sources.

- Velocity – rapid collection and analysis of data

For emerging services in the transport domain (e.g., automation), real-time data processing is a required. For example, all stages from data collection to traveller information services are performed in real-time. However, while 'velocity' is generally seen as a pre-condition in the broader big data context, in the transport sector there are services based on big data that do not require real-time processing. This is particualrly the case for non-safety-critical services, such as in optimising maintenance services wherein the planning cycles do not require rapid data processing.

- Variety – data is in varied format and usually from various sources

Big data comes from various sources and therefore most likely would be in different formats. While the variety of data is typical, there are examples in the road network operations wherein sigfnicant amounts of data can come from only one source, e.g., floating car data, and they may be collected in a consistent format.

Aside from these three major V's, big data concepts describe following 2 V's that are of minor but relevant importance to road network operations:

- Veracity – datasets may contain inaccurate/unreliable data

Especially in safety critical services, inaccurate data is highly undesirable. Therefore, when dealing with big data road network operators need to take precaution against use of potentially erroneour data.

- Variability – data quality changes (e.g. over time)

Inconsistency in the data quality is undesirable for road network operator applications. Transport operators and service providers would need to take carefully calculated determinations of risks as a result of inconsistency in the quality of big data.

Data scientists consider more V´s to describe big data even further e.g., visualisation, value, volatility, validity. While these are pertinent in the broader context of big data across all sectors, these additional attributes are not expected to be significantly relevant in the context of this report, hence they were not considered in detail.

Big data applications are becoming more developed. Meanwhile a series of initial applications are already in operation that could serve as a pathway from 'small data' applications to 'big data' applications. Generally, these transitional applications may be referred to as 'large data' applications. This report is restricted to strict definitions of big data, but it also covers large data applications, which do not have feature all the attributes of big data described earlier (i.e. 5 V´s). To simplify the terminologies in this report for an easier read, the term big data in this report refers to both big data and large data.

2.2. FROM DATA TO SERVICES

Big data in transport is an enabler that will supply the travelling public with solutions to better meet their mobility needs, as well as support the network operators whose goal is to improve safety, efficiency, accessibility, as well as the reduce the environmental impact of the whole transport system. Big data has the potential to also support network operators, especially in the following:

- implementation of new services, including Cooperative Connected and Automated Mobility (CCAM), Mobility as a Service (MaaS), and new tolling schemes
- day-to-day transport network operations, including traffic management and activation of traffic management plans
- accurate traveller information services
- planning, design and maintenance
- budgetting and procurement, as well as, contract management.

Figure 1 illustrates the main steps of the big data process, from generation to information or even services.

Figure 1: Big data process (summary view)

1. **Data sources:** Data sources does not only comprise of traditional sensor data e.g. traffic data collection; but other non-traditional sources are increasingly gain importance including positioning data derived from cell phones, navigation systems and cars, free floating bikes, Electronic toll collection (ETC) readers, and on-board units on trucks. Social media data is being used to some degree, but it is expected that they could provide further useful information in the future. A close look on the different sources will be provided in the later chapters of this report.
2. **Data Collection:** Road network operators will aim to acquire data wherever and whenever available for analysis. One constraint is the capacity for data storage, however there are technologies that can address the issue of storage capacity. For example, edge computing (i.e. data handling at the sensor) is an efficient way of reducing data size for in the cloud or data lake.
3. **Data Preparation/Curation:** Because data is sourced from various sources, a key task in the big data cycle is to prepare the data such that it is in a format to facilitate analysis in the next

step. Data curation is the organisation of data, inclduing determination of what information is saved and for how long.

4. **Analytics:** Processing of data using big data analytics technology, such as data mining, is a key element of the big data cycle.
5. **Visualisation:** The preparation of new knowledge as a result of analytics is often under-estimated. Visualisation refers to presentation of analysis results in a organised and understandable way, which would facilitate understanding and decision-making.
6. **Access:** The resulting information is the basis for services, which often generates new data for further processing. Hence, it is of high importance to provide information to as many users as possible. Moreover, in order to gain as much added value as possible, the data should to be as open as possible so that new applications may be developed.
7. **Information, Valuable:** The most important outcome of the process is the added value to society, which often is a result of new or improved services. Furthermore, one other result of the big data cycle is new information, which is often not expected when the data is collected and analysed.

2.3. THE BIG DATA CYCLE

Cooperation between various stakeholders, including public and private entities, is key to maximise generation of knowledge (see Figure 2). This cooperation is especially needed because the big data process is dependent on data sources from different data owners, which is a core element for the resultant end-user services. Big data generated by public and private stakeholders are used to generate information in the big data process (see chapter 1.2). The collected information forms the basis for the generation of enhanced transport services.

In addition to internally generated data, data from 3rd party sensors contribute to validate or assess the impact of services provided. As an example, roadside sensors as well as floating car data (FCD) can be used to measure the influence of variable speed limit signs to traffic flow and speed. The measurements is again data which is further input of the big data process; and, it provides information if the service meets its objective and if adjustments are needed.

Through big data techniques, the cycle of data measurement - evaluation – service becomes faster and more accurate over time and this leads to potential new mobility services, such as CCAM and MaaS.

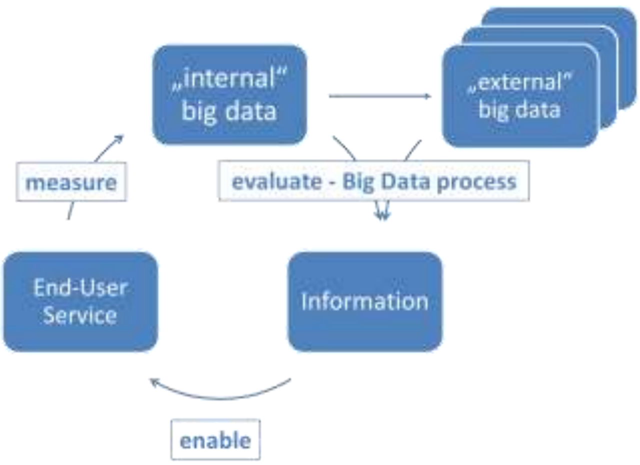


Figure 2: The big data cycle

The cooperation of stakeholders involves a large number of diverse parties, and the key stakeholders are listed in Table 1.

Table 1 – key stakeholders and their involvement in the big data Service

STAKEHOLDER	DATA OWNER	KNOWLEDGE GENERATION	SERVICE PROVISION	SERVICE USAGE	INFLUENCE ON SERVICE
Individual					
Travellers	X			X	X
Public					
Transport operators	X	(X)	X	X	X
Toll operators	X	(X)		X	
Emergency/rescue services	(X)			X	
Policy Makers				X	X
Private					
Telecom operators	X	X	X	X	
OEMs	X	X	X	X	X
Automobile clubs	(X)	(X)	X	X	X
Fleet operators	X			X	
B2B service providers		X	X		X
Data Warehouses	X			X	X
Data Platform operators (e.g., Google)	X	X	X	X	X
Media	(X)			X	X

The services are for the transport users. But in most cases transport users are not only the recipient of services, but they are data owners as well. Entering origin/destination data in a smartphone is an example of this data. Moreover, travelling with a cellular device (which might be a smartphone or in future a connected vehicle) results in movement data, which is valuable to road network operators in identifying possible disruptions in traffic patterns, which is indicative of incidents.

Like transport users, public sector stakeholders generate data as a result of the services they provide. Transport operators have an incentive to generate and maintain accurate traffic data to support their traffic management activities as well as end-user services. Without accurate traffic data, which typically include speed limits, movement restrictions or closures of parts of the network, the effective and efficient management of the network and the provision of qualitative end-user services would not be possible.

The private sector currently has a major role in the processing of big data. For example, telecommunication operators have tracking data on cellular devices operating on their network. The tracking data can be utilised to generate and visualise traffic flow on the transport network in real-time. Private firms also purchase data from various sources for various transport applications. For example, the Shenzhen Urban Transport Planning Center (SUTPC) in Shenzhen (China) collects data from 3rd parties (Telco, ETC, Mobike, subway operators, etc.) for urban planning purpose.

The involvement of private sector points to a question that if in the future the know-how of processing big data should be an internal competence of transport operators, or if relying on external parties will be more appropriate. In either case, it is evident that there is potential benefit in big data processing; hence it is worthwhile to further consider big data in more detail.

3. BIG DATA ENABLING FUTURE MOBILITY

3.1. RELATION OF BIG DATA WITH OTHER TECHNOLOGY TRENDS

Several promising technologies are on the horizon that are related to big data. Although these technologies are not a core aspect of this report, it is worthwhile to briefly review them and to examine the role of big data in these technologies:

Internet of Things (IoT)

IoT is the concept of connecting any electronic device to the Internet (and/or to each other). This includes almost any device or equipment such as cell phones, coffee makers, washing machines, headphones, lamps, wearable devices and others. This also applies to components of machines, for example the jet engine of an airplane or the drill of an oil rig.

IoT is a significant source of data and hence, an important topic to considering big data. Conversely, data analytics is utilised to generate benefits from IoT, which results in a symbiotic relationship between big data and IoT. Big data analysis is gets its input data from IoT, and at the same time, big data analysis enables the benefits from IoT.

Edge Computing

IoT generates a large amount of data thereby IoT-related technology is subjected to limitations in network latency and bandwidth when transmitting from data to the cloud, as well as, limitations in cloud capacity (economically or technically). Edge computing enables sensors to process data directly at the source rather than sending the data unprocessed to the cloud. Data mining can be done within the sensor making applications such as traffic lights be more responsive and intelligent in optimising operations through computations being done onsite.

Fog computing is related to edge computing, wherein data processing takes place somewhere in between the sensor and the cloud.

Artificial Intelligence (AI)

Artificial Intelligence have been rapidly improving and are finding real-world applications such as virtual assistants. AI goes hand-in-hand with machine learning and developments in machine learning is further advancing AI applications. Machine learning requires data to work, and therefore has close links with big data.

Smart concepts

Smart concepts link modern technology in the fields of energy, mobility, spatial planning, administration as well as communication with the aim to improve quality of life. Smart concepts are often used in urban environments to make them more sustainable, which are also called smart cities. Smart cities are integrally linked with IoT and consequently to big data, as noted above.

Building Information Modelling (BIM)

BIM involves the digital representation of a facility. BIM thereby generates asset data when the asset is designed. The data includes maintenance and service functions, which enables infrastructure owners to develop a clear picture about the state of their assets, how they need to be managed, and what resources might be needed to maintain their capability. So far, BIM is mainly focused on the construction and operation of buildings. However, road network operators are also considering using BIM for their networks. Data generated from the setup of the facility with BIM functionality generate a significant amount of data, which has applications in smart cities. An example related to transport applications is car parks collecting real-time data on the number of vehicles entering and exiting the car park. Parking traffic data has the potential to be applied by parking space prediction models for parking advisory services.

Data Warehouse

In a traditional data warehouse, the data handling process (i.e. collection, preparation, storage, and analysis) is similar to a big data process. However, the big data process differs from traditional data warehouses in the ordering (sequence) of the data handling for big data, for example:

- big data volume system: data is stored immediately in raw form before data preparation
- big data management: collection, preparation, and analytics occurs in real-time, and may include data aggregation prior to storage
- data warehouse: data storage occurs only after data preparation and aggregation.

3.2. INTEGRATED SERVICES ENABLED BY BIG DATA

In addition to the maintenance and operation of the network, transport network operators also inform travellers about the status of the transport network. Traditionally, Traveller Information Services are provided to travellers via radio, roadside Variable Message Signs, public transport information displays, internet, mobile phone applications, and in-vehicle systems. It is important to deliver a high-quality service in terms of reliability, accuracy and timeliness of the information provided.

While in the past communication media (e.g., Variable Message Signs) have been used to inform travellers, future services are looking into integrated services, wherein information from infrastructure operators is integrated into end-user services. In this case, the service provider integrates information, data and services from different stakeholders into a singular tailored and personalised end-user service. The task of the infrastructure operator hereby is to provide reliable safety and efficiency related information as input to the integrated end-user service.

Two examples of such integrated services are CCAM and MaaS, and they are examined in more detail as follows.

3.2.1 Cooperative, Connected, Automated Mobility (CCAM)

CCAM encompasses connected vehicles (also referred to as cooperative systems or C-ITS) and automated road transport (including automated and autonomous vehicles). Both aspects are covered in detail by other PIARC task forces, namely ‘TF B.1 Road Design and Infrastructure for Innovative Transport Solutions’ and ‘TF B.2 Automated vehicles: challenges and opportunities for road operators and road authorities’. Big data, IoT and AI are key enablers of CCAM including facilitating new comprehensive road network service as shown in Figure 3.

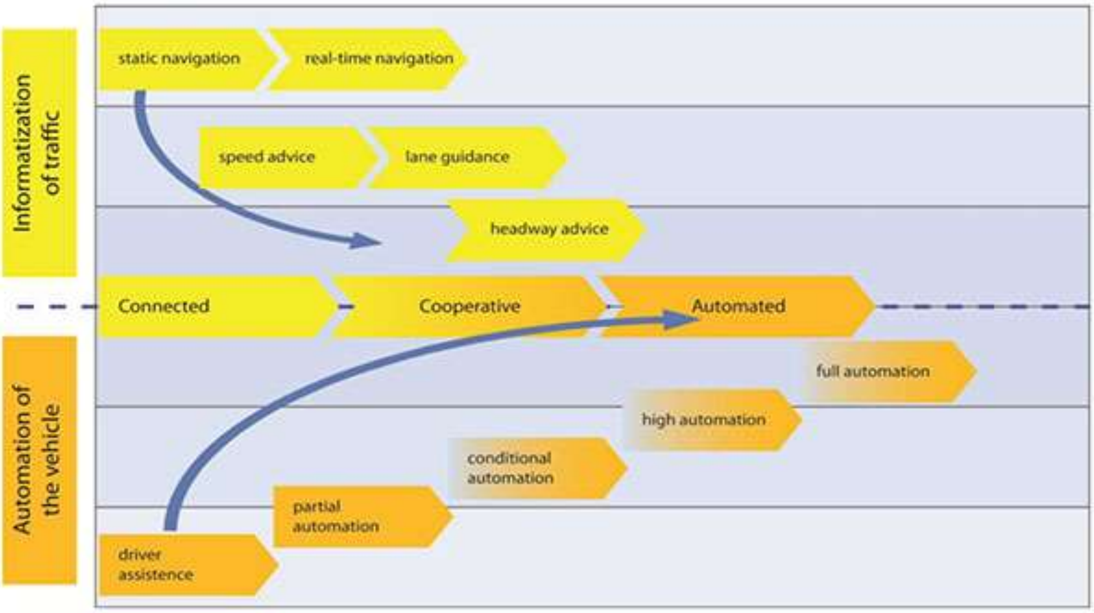


Figure 3: How automation and information will be merged towards cooperative connected automated mobility [1]

The core element of CCAM that has the most relevance to big data is related to its cooperative aspects. Future road transport services would be based on cooperation of different stakeholders. In the context of a fully self-driving car, i.e. one that is able to drive from origin to destination without driver intervention (or even without a driver), it is essential the data is available to ensure a safe and secure trip.

The next generation of cars would be able to utilise sensors to gather information about its driving environment to be able to operate autonomously, but sensors are only able to do for its environs that are of close proximity. Current sensors have limitations and they have a range of approximately 300 meters around the vehicle. There are therefore risks related to driving hazards beyond what the sensors can detect. In these scenarios, communication (or connectivity) between different stakeholders becomes important. Vehicles would be able to communicate with other vehicles (even between vehicles from different vendors). Moreover, road network operators would also be able to inform vehicles about the road status, such as driving restrictions, roadworks, traffic light status, and traffic management strategies. This information is important in rural highway operation, and is perhaps even more critical in urban areas, where there are vulnerable road users.

In CCAM, big data processing of data from various sources (e.g., weather data, data from event-organisers) in real-time is critical to ensure a safe trip for travellers.

Figure 4: Truck platooning becoming a future challenge for road network operators [2]

Big data processing in CCAM is not only important for autonomous driving, but it is also valuable for road network operations. Big data would improve the performance of traditional tasks of a road network operator. Big data, in conjunction with CCAM, would also enable new types of service available to network operations. One such service is vehicle platooning on highways, wherein two or more vehicles are closely following each other while they are supported by automated systems. During platooning and when coming to motorway exits and entrances, one important manoeuvre is managing dispersing or breaking up the platoon to enable other vehicles to merge or diverge which otherwise would be difficult with tightly spaced vehicles in a platoon occupying one lane. To disperse the platoon, the following data would be needed:

- presence of a platoon and its attrbutes including speed, exact position (including used lane), direction, etc.
- geometry of the highway including entrance and exit ramps
- presence of cars that plan to leave or enter the highway.
- environmental information on the general traffic flow characteristics on the highway, weather conditions
- status of the platooning vehicles (e.g. technical problems).

The above data from various sources would need to be collected, analysed and a traffic management strategy needs to be determined and communicated in order to disperse the platoon. One example of a dispersion platoon strategy might be ‘dissolve the car-platoon between the 3rd and 4th vehicle’; stop the entering vehicle to give way for the car platoon; or, slow down the exiting vehicle. In any of the three example cases, decisions and strategies will be based on real-time data processing and that the data needs to be perfectly reliable, robust and accurate.

3.2.2. Mobility as a Service (MaaS)

New generation multimodal services are commonly described as MaaS. MaaS is an end-user driven service whose goal is to ensure that a traveller reaches their destination by whatever means available and in manner, that best suits the traveller. The MaaS Service Operator provides tailor-made mobility services utilizing all kinds of transport modes and covering all aspects of organising the trip from planning, booking and payments. In its advanced form, MaaS could be provided services to travellers under monthly mobility service packages, which allow travellers to use various transport modes to best suit their individual needs.

In the MaaS environment, it is expected that a traveller will no longer need individual tickets for each transport modes they take, instead a combined MaaS ticket will allow travellers to use a multitude of modes, including public transport, car, shared bike or taxi. Managing the travel itinerary including travel information, booking and payment of fares or fees is centrally organised by the MaaS Service Operator.

One key enabling technology for MaaS is big data processing. The MaaS Service Operator needs to get access to static as well as real-time data of transport modes (either privately or publicly organised). Static data (i.e. timetables, transport network, fare prices, etc.) are combined with real-time data (i.e. traffic status, delays, occupation rate of taxis, etc.); and, the fusion of these data enables a MaaS Service Operator to provide tailored services for individuals. MaaS would also need to link and integrate various services (e.g. information or ticketing services of transport operators) to be able to provide sufficient and sustainable MaaS service.

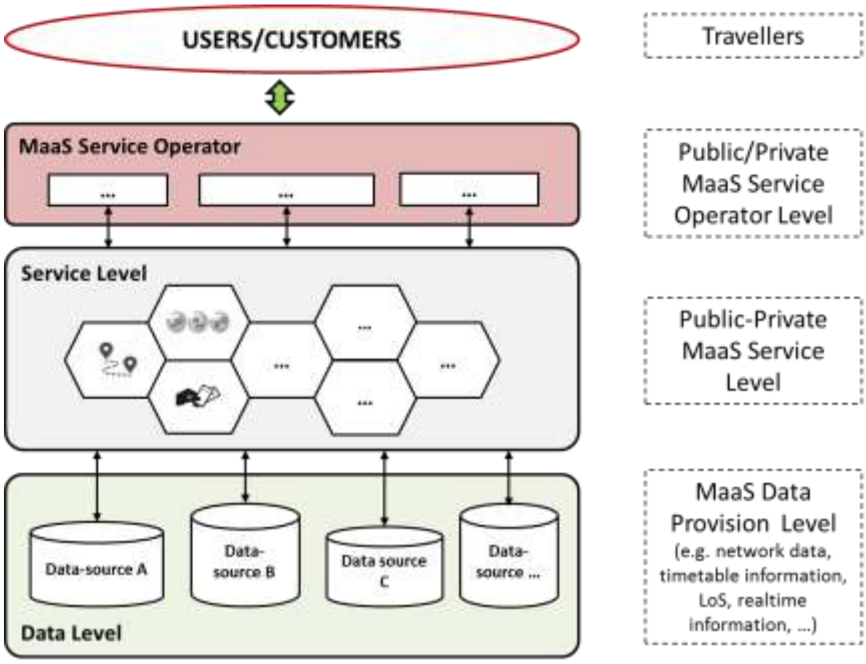


Figure 5: The MaaS Ecosystem [3]

A possible MaaS ecosystem, as shown in Figure 5 and it would include big data processing at different levels, including:

- public-private MaaS service level, where e.g., routing services or payment services are generated
- public or private MaaS Service Operator level, where data and services are integrated into one tailored MaaS end-user service.

In the MaaS ecosystem, it is expected that the data flow will not only be from the transport operator via the MaaS Service Operator to the end-user, but vice versa as well. Hence, end-users through their mobility requests will generate data on their mobility needs and gaps. For example, requests for multi-stop transport service requiring multiple intermediate transfers is an indication that there may be a need to improve the available services, such as a new service route. The need for additional public transport lines as well as the need for improved transport networks can therefore be assessed and determined from the data. As a consequence, MaaS will support transport network operators in improving the transport network in based on users´ requirements.

3.3. THREATS

New developments tend to be accompanied by expectations, which can vary but often the expectations are that these developments would lead to early and significant returns. Policy makers and decision makers tend to expect quick wins in order to rationalise budgetary outlays. If big data processing is not able to meet the unrealistically expected return then frustration might lead to premature cancelation of budgets for big data. It is therefore of utmost importance to clearly set and agree on expectations; and, to start slowly and to progress at a suitable rate. In order to do so, new expertise is needed in technical fields that are in most cases new to road operations. Traditionally, road network operators are focused on civil engineering, hence, a change in thinking is often required which is sometimes only possible over a certain period of time. Furthermore, IT and related applications will play an increasingly important role in all core businesses of road network operation. Hence, investment in personnel and equipment are required.

The development of technical expertise is important as data analysis has high risks in arriving at wrong conclusions due to variety of reasons, including limitations in data sources. For instance, mobile phone-sourced data quality depends on the context of use (e.g. time of day, type of user and speed); hence, data analysis without considering these factors could lead to potentially erroneous interpretations. In the worst cases, errors could cause long-term damage to the confidence in big data results. Therefore, big data processing needs to apply a well-developed data quality assessment. Otherwise, there is a risk of low level of acceptance with end users and as a consequence negatively impact the reputation of data providers.

Once big data has been set-up as part of day-to-day operation of the road network, the application of machine learning technologies is applied to further improve performance of the network. The application of machine learning should be carefully considered, wherein the correct algorithms are implemented, and quality checks are constantly ensured.

Another key consideration is that a single source of truth is maintained, as data and services are utilised by the public and for business-to-business applications. It needs ton ensured that third parties do not degrade the quality of the data and that it is used in pursuance of its transport service obligations.

Violation of privacy rules also poses another challenge to be met by big data. Legal departments of road network operators need to prepare for these increasingly important questions.

Finally, security issues also pose another highly important challenge. For instance, probe data applications might be manipulated by malicious parties by masquerading their devices as authorised devices which allows them to manipulate data inputs. Public and private providers will have to take adequate measures to avoid security breaches. Frequent security failures will consequently lead to unreliable services which will result in confusion amongst drivers, reduced participation in using the service, decreased number of users providing data, and decline in accuracy of probe data.

4. IT IS ALL ABOUT DATA

4.1. WHAT IS DATA

Data is factual information (such as measurements or statistics) which is used as a basis for reasoning, discussion, or calculation. In digital form, data can be transmitted or processed. Data is an information output of a sensing device that includes both useful and irrelevant or redundant information and must be processed to be meaningful. In addition, data is also generated by social media and social media has a contextual nature.

4.2. DATA EVOLUTION

Data is generated from a variety of different sources and each data source contributes to data analysis and data processing resulting in the generation of valuable information. Typically, the data collection process (refer as well to chapter 1.2) starts with single or multiple source where data comes from. During the data collection phase, it is important to establish the data format. The various formats of data are described as follows:

Digital data

Digital data is information that has been captured through a digital medium, e.g., a computer, smartphone app, or sensor with digital signal processing.

Analog data

Analog data is information that is available as a picture, video and other similar formats. For big data applications, analogue data is required to be converted to digital formats through the use of sensors, such as cameras, voice recording, and digital assistant. The rapid advances in digital technology has accelerated the rate at which traditionally analogue data is being converted or captured through digital mediums (e.g. digital signal processing).

A key step in the data collection process is the storage of data sets followed by data preparation for further processing. It is advisable to store data in a format that is consistent with existing standards or to establish a standardised format so that subsequent analytics and visualisation tasks can be done efficiently and accurately. By doing so data becomes much easier and allows for much better and more accurate outputs. For example, spatial data are mainly exchanged by using the GDF (Geographical Data Format) even though are internally processed using proprietary formats.

4.3. DATA AND TIMELINE

Data sources can be acquired for various purposes and in various forms. One major attribute of data is time, i.e. when and at what interval was the data collected). Each time attribute has its value depending on its applications. Data can be categorised based on their time attribute as follows:

- Static data
- Real-time data
- Dynamic data
- Historical data

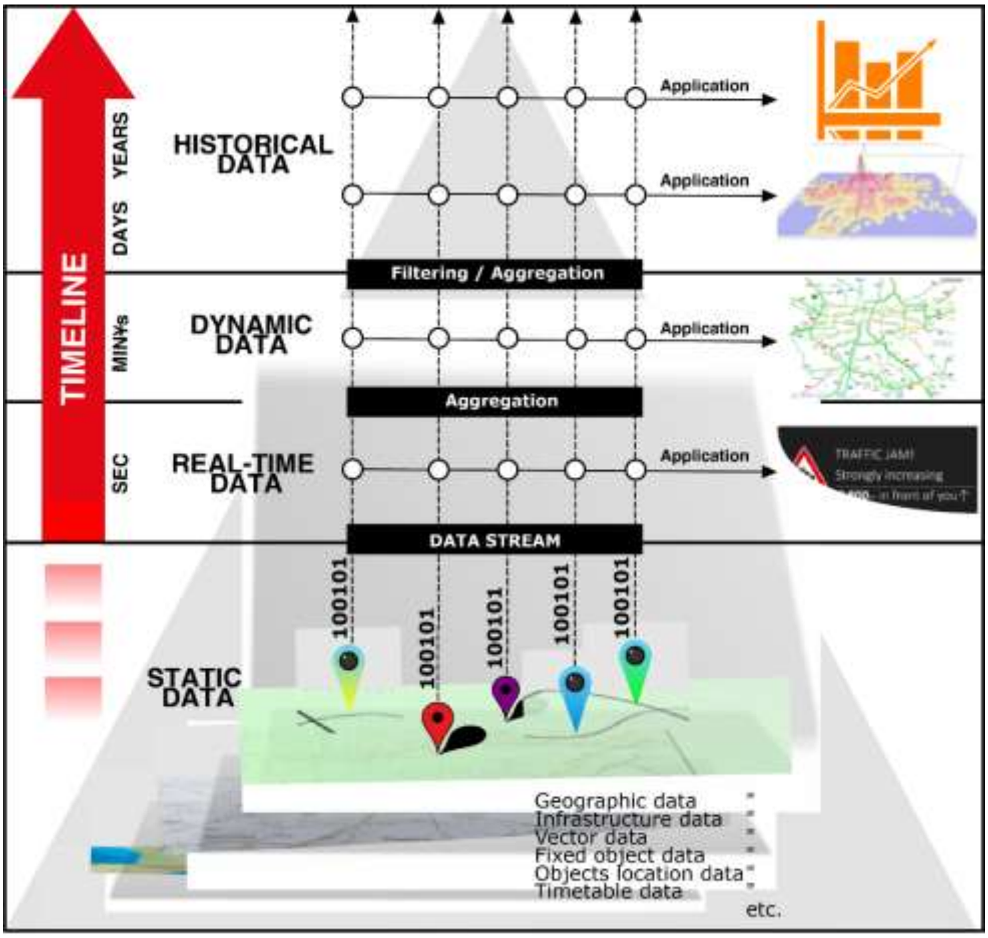


Figure 6: Data and Timeline

4.3.1. Static Data

Static data does not change, or it changes in very long intervals, hence static data is considered to be self-contained or controlled. In the case of transport monitoring, static data is primarily the transport network data, i.e. map of highway segments. The basic transport network data layer comprises of geospatial data. Nowadays, big data analytics of geospatial data is growing and it has enabled users to effectively and efficiently analyse large amounts of geospatial data.

Road network operators collect large amounts of data to provide some visibility of their network. Data sets can be refined by regular updates and available dynamic data could be displayed and overlaid on top of the static data. Static big data includes the following:

Technical data of roads

Technical data includes cross-section of a road (lanes, roadside, etc.), surface type, safety features (barriers, etc.), road objects (ford, bridge, underpass, tunnel, railway crossing), roadside objects (curb, fence, wall etc.), traffic signs (gantry, plate), lane marking, and others.

Road infrastructure objects with locations

Road infrastructure objects include rest areas; parking lots; park & ride; weigh stations; telematics equipment; toll gates; petrol , LPG/CNG filling and EV charging stations; surveillance and monitoring equipment; signalling and control equipment; etc.

Traffic regulations

Traffic regulations include restrictions (static speed restrictions and advisories); traffic restrictions; static signs; permanent directional signs; etc.

Public transport

Public transport and multimodal transport data include timetable and facilities i.e. bus stops, terminals, platforms, sidewalks, parking lots, escalators or lifts, bike-sharing docking stations, car-sharing stations, etc.

Static data is very important and without suitable static data, all dynamic processes are limited; hence, collecting, analysing and visualising static data needs to be carefully considered.

The current trend in static data is to develop a precise and detailed representation of the entire physical infrastructure in digital form. In particular, the goal is to create high-definition (HD) maps that include accurate lane-level geometry, placement of all traffic control signs and advisories, traffic controls at intersection junctions, and major street furniture. HD maps are one of the core data requirements to support automated driving and connected vehicle functions in the future.

4.3.2. Real-time data

Real-time data combines streaming data or event data. Event data is collected from sensors or from moving objects (e.g. vehicles, trucks, and personal devices) or it could also come from web feeds. Real-time data involves large amounts and varied types of data going into servers. Real-time data is constantly changing. Due to its size, speed and flow variability, it is problematic and difficult to store and analyse it in its raw form. However, real-time data is very valuable because it provides information of real-time events in a digital signal. Hence, for real-time data processing, it is necessary to develop and implement optimised analytical techniques, because processing is done in a short amount of time and with limited resources (e.g. memory) to produce accurate real-time results for traffic control applications.

4.3.3. Dynamic data

Dynamic data is generated when real-time data from various moving objects is evaluated and aggregated at short time intervals. Dynamic data is therefore based on real-time data collection. Dynamic data includes average speed on lanes, number of vehicles by class, headway, average weight of vehicles, and others. The analysis of dynamic data results in useful information such as congestion and delay.

Dynamic data is also used in the public transport sector, e.g. determining delay on lines and occupancy of vehicles. The data may be collected by on-board detectors to estimate the number of passengers or by sensors that can detect contactless cards carried by passengers entering and exiting vehicles.

4.3.4. Historical Data

Historical data are filtered and aggregated dynamic data which are stored data for an adequately long period of time. With historical data, systems can become robust and reliable as they can be used to improve predictive models used as a core element in traffic operations. With improvements in cloud storage solutions, data storage space limitations are no longer constraints and the data can also be accessed easily.

4.4. DATA SOURCES

Traffic operators have a variety of tools available to monitor the current state of the transport infrastructure. Sensing technologies available fall mainly into two broad categories: infrastructure based and non-infrastructure-based sensors.

To date, infrastructure-based systems are dominant and there is a vast range of sensor types available, ranging in complexity from pneumatic tubes that are run across the road to laser vehicle profiling systems. In addition, sensing systems are increasingly being used to monitor traffic across the whole area and not only in a single road traffic profile.

4.4.1. Infrastructure-based traffic detectors for information on the point

Sensors have provided traffic management for over 35 years with information on traffic volume, incidents, device malfunctions, speeds, and environmental conditions. Road network operators will continue to deploy more infrastructure-based sensors and other devices for traffic and transportation management, thus continuing to increase the load on existing processing, storage, and analysis systems in traffic management centres (TMC).

Commonly used infrastructure-based detectors are:

- **Inductive loop detectors** consist of one or more loops of wire in the road surface whose inductance is changed by the presence or passing of metallic objects.
- **Magnetic sensors** are mounted under the road pavement and they detect a change in the earth’s magnetic field caused by the passage or presence of a vehicle. These sensors are typically battery powered and connected to a roadside equipment wirelessly. Magnetometer-based systems can count vehicles and determine occupancy at the same level of accuracy as inductive loops and they are easier to install and their installation causes less pavement damage.
- **Video image processors** are mounted over the road. These systems have the potential for generating the greatest amount of information but are probably the least reliable. Video systems can collect the same data as loop detectors, and they can also collect speed, queue length and vehicle classification directly. When equipped with automatic number plate recognition (ANPR) capabilities, pairs of cameras are capable of collecting travel time and origin-destination (OD).
- **Microwave radar sensors** are mounted over the road and they measure vehicle troughput on a specific road section.
- **Infrared sensors** are categorised into active and passive infrared sensors, based on their power supply, and they are used for traffic flow monitoring applications. Both types can can count and classify vehicle types. Infrared sensors are mounted over the road.
- **Laser radar sensors** can scan infrared beams over one or two lanes or use multiple laser diode sources to emit a number of fixed beams that cover the desired lane width. They can measure speed and axle count.
- **Acoustic sensors** are microphones (or microphone arrays) that listen for traffic noise. Some of these types of detectors can count vehicles, but their accuracy is limited.

- **Ultrasonic sensors** transmit high-powered sound waves above the range of human hearing and they can detect the reflections from objects (similar to radar).
- **Piezoelectric or optical fibre technology** is mainly used for weigh-in-motion systems that measure vehicle mass and axle counts as vehicles pass over them at traffic speeds.

A widely adopted infrastructure-based vehicle detection system is the inductive loop. In its basic implementation the inductive loop yields vehicle count, vehicle speed, occupancy (effectively point density) and vehicle class. Even though inductive loops are widely deployed, they are unable to track traffic origin-destination (OD) behaviour and, consequently, travel time. Inductive loops are usually connected through wired or wireless communications and form sensor networks.

Table 2 – Infrastructure-based sensors and their attributes

DEVICE	SPEED	CLASS	OCCU-PANCY	HEIGHT	WEIGHT	MULTI-LANE	LIMITED BY WEATHER	COST
INDUCTIVE LOOP DETECTORS	yes (in pair)	yes	yes	no	no	no (each lane)	no	average
MAGNETIC SENSORS	yes (in pair)	yes (in pair)	yes (precise)	no	no	no (each lane)	yes (ice on surface)	low
VIDEO IMAGE PROCESSORS	yes (limited)	yes	yes	yes (limited)	no	yes	yes	Average to high
MICROWAVE RADAR SENSORS	yes (precise)	yes	no	no	no	yes	no	Low to average
INFRARED (IR) SENSORS	yes (limited)	yes (limited)	yes	yes (limited)	no	no	yes	low
LASER RADAR SENSORS	yes (precise)	yes (precise)	yes	yes	no	possible	yes (heavy rain, snow)	high
ACOUSTIC SENSORS	yes (limited)	no	no	no	no	no	yes	low
ULTRASONIC SENSORS	yes (limited)	yes (limited)	yes	yes	no	no	no	low
PIEZO-ELECTRIC (WIM)	yes	yes (axles)	no	no	yes	no (each lane)	no	high

4.4.2. Sensors for information along road stretches

To get data along road stretches between two points along the road, in-vehicle data is necessary. They are also commonly known as 'A-to-B systems'. Examples are described as follows:

Floating Car Data

The best-known vehicle-based sensing system is floating car data (FCD), wherein vehicle movements are monitored and used for speed information. The basis for FCD is technologies that allow a vehicle to detect its position via a satellite and to transmit its position at a sufficiently high sample rate (typically every several seconds to several minutes) to a FCD centre. There the data is anonymised, and trajectories generated are used to determine travel speed.

Usually, FCD is from commercial fleet vehicles (e.g. taxi) whose vehicles are equipped to provide data whenever the vehicle is in motion. Vehicles act as sensors for real-time traffic information by providing information on their speed, position and direction of travel. FCD can be used as a primary source of data for road networks that have no or limited infrastructure-based sensors; or, as a secondary source of data to complement infrastructure-based data. FCD tends to be reliable if the penetration rate of FCD equipped vehicles in the traffic stream is high enough, i.e. a penetration rate of 15 to 20 percent is sufficient. One limitation of FCD is the data ownership, as the data is owned by commercial entities, thereby negotiation on access and usage tends to be an issue.

Portable devices

Portable devices are mobile devices that generate and transmit data along a trip. Similar to FCD, the movement of the portable device are collected and analysed. Usually portable devices are tracked by service providers who aggregate the information from several other tracked devices. The data is analysed, and the information is provided back to the traveller for a variety of location-based services. The same information may also be shared with other stakeholders including road network operators. This type of data is usually collected by private companies (e.g., Google, Waze, TomTom, DiDi, Kakao) using their mobile applications. Travelers allow these companies to collect and use the information about their trips as the application provides navigation services. The disadvantage of this data source for road network operators is that it does not provide a guaranteed level of data quality. Data generation is dependent on user needs as well as service usage at the time of data collection.

Point-based tracking

Point-based tracking utilises roadside sensors communicating with vehicle-based devices to collect data. Examples of these devices include Dedicated Short-Range Communications (DSRC), WiFi, Bluetooth, ANPR, or tag toll readers. DSRC is technically a modification of the consumer WiFi, and they are designed to operate efficiently in a (physically) high-speed mobile environment (e.g. 200 km/h speeds) and support low-latency applications such as in vehicle safety (e.g. vehicle collision mitigation).

In public transport, data from smart card systems are collected and processed when passengers validating their travel cards. The data on smart card transactions can be used to optimise the network including to identify new routes and lines and to modify timetables. For operators, it is advisable to set up the smart card ticketing system such that passengers validate their travel cards not only when they board but also when they disembark the vehicle.

Connected vehicles

A specific case of point-based tracking is the connected vehicle. Connected vehicles can communicate their status directly to other vehicles, other road users, and roadside systems so that every vehicle on the road is aware of where other nearby vehicles and travellers are located. Connected vehicles can identify threats, hazards, and delays on the roadway and relay the information to the road network operator who could use the information to provide drivers with alerts, warnings, and real-time information. Connected vehicles may use a variety of communication media for data exchange, including DSRC and cellular communication networks.

Drones

An innovative way to collect data from moving objects is the use of drones. Traffic data collected from drones can potentially provide a rich new source of data on traffic movements. Drones can be equipped with high-resolution cameras, high capacity data links, video image processing, and automated navigation systems with flight times of several hours. However, drone applications is regulated in some countries by authorities on civil aviation; nonetheless, their potential in traffic management is evident in cases when they are allowed to operate by the authorities.

Dash cams

Dash cams are typically used in incident management on highways. Dash cams installed on patrol or incident response vehicles and combined with on-board telematics equipment they can provide traffic management centres (TMC) with live on-site video visibility of the road. This information can enable the operators to implement mitigation measures faster and more precisely when needed. TMC operators can access the remote video stream on a web platform to allow them better situational knowledge hence they can better address the impacts of an incident including to ensure the arrival of the various response teams. The video data can also be used as evidence for insurance purposes.

4.4.3. Users-based data

Mobile devices (i.e. smartphones and tablets) have become ubiquitous these days. Smartphones can be used to track the trajectories of the person carrying them. Especially since the capacity of batteries and the efficiency of application processors have improved significantly it has become possible to record the location of a person with a high level of frequency. Mobile devices employ either an installed application or a subscriber identity module (SIM) card for detection

Data collection through the installed application

Global Navigation Satellite System (GNSS)-enabled smartphones often automatically collect location data through installed applications. Data generated and transmitted by the application users could include location, trip characteristics (e.g. speed), or transport mode (e.g. car, transit, walk or bike). Data from mobile devices has the potential to provide information that road agencies can use to improve traffic management, as well as, urban planning.

Data collection through the SIM card

There a significant number of mobile wireless devices that register at cell towers or Base Station Subsystems (BSS) while traveling along its proximity. The pattern of movements of these devices can contain key information on the performance of the transport network. There are two main methods for locating mobile phones through the telecommunication infrastructure, i.e. active and passive mobile positioning.

Active mobile positioning is a process where a mobile phone is tracked in real-time using measurements from the network. This can be achieved through trilateration through active ping for the phone in the network, which is similar to making a phone call without it actually taking place.

Passive mobile positioning is extracting data on the historical location off the phone from the log files of telecommunication operators. This method allows for a longitudinal view of all subscribers. Any smartphone that is turned on works as a sensor that is being carried around by individuals, and the smartphone is able to collect data on movements and activities.

Table 3 – Non-infrastructure-based sensors and their attributes

DEVICE	SENSOR LOCATION	REAL TIME TRACKING	DATA	TRAFFIC VOLUME	VEHICLE CLASS	SEGMENT TRAVEL TIME	COST	VEHICLE SUB-SYSTEM STATUS
VEHICLE-BASED								
On-board units (i.e. FCD)	GPS/GSM (in vehicle integrated)	yes	Anonymised ID, Location, Speed, Direction, Timestamp	no	yes (depends on registry)	yes	average	Wipers status, brake, acceleration-traction control, ABS
Portable navigation devices (Crowd-sourcing)	GPS/GSM (mobile device)	yes (if the application is ON)	device ID, Location, Speed, Direction, Timestamp	no	no	yes	low	no
POINT-BASED TRACKING								
Bluetooth, WiFi	Reader x Tag	no	MAC, ID detector, Timestamp	no	no	yes	low	no
Toll tag readers	Reader x Tag	no	tag ID, ID detector, Timestamp	yes (toll collectio n system)	yes (depends on registry)	yes	high	no
ANPR Cameras	Camera x Licence plate	no	ID detector, License plate ID, Timestamp	yes	yes (limited - privacy issues)	yes	average	no
Connected vehicles - C-ITS	RSU x RVU	yes (on equipped infrastruct ure)	Anonymised ID, Location, Speed, Direction, Timestamp	no	yes	yes	high	Wipers status, brake, acceleration traction control, ABS
USER-BASED								
By Installed application	GPS/GSM (mobile device)	yes (if the application is ON)	device ID, Location, Speed, Direction, Timestamp	no	no	yes	low	Acceleration, vibrations, etc.
SIM card detection	SIM card (phone, tablet)	no	device ID, Cell ID, Timestamp	no	no	yes (limited)	average	no

*	Traffic Volume describes the possibility of counting vehicles. (YES= The technology is able to count vehicles; NO= The technology is not able to count vehicles).
**	Vehicle Class describes whether it is possible to classify the vehicles. (YES= The technology is able to classify vehicles (The classification of vehicles e.g., individual vehicle, Truck, Bus, etc.); NO= The technology is not able to classify vehicles)

4.4.4. Other data sources

There a variety of other sources of data which can also be used by transport operators. Among the most important are as follows:

- Meteorological sensors: wind speed and direction, precipitation, humidity, temperature, visibility, etc.
- Air polution sensors: NOx, CO2, particles, etc.
- Noise sensors: level of noise
- Accident data: crash or accident data are typically collected by the national police agencies. Crash data is usually accessible through databases for easy analysis and annual reporting.
- Natural hazards monitoring: includes data on earthquakes, landslides, water level, movement of snow masses, object monitoring (bridge movement, road surface movements, etc.)
- Social network data: a social network related big data could be derived from internet-based social media/network tools. Social media is an internet-based system that disseminates information to others.

The above data sources are used in combination with other data sources. The fusion of data analysis facilitates a comprehensive understanding of the network, which can be used for a variety of purposes.

4.5. METADATA

Big data is a collection of large amounts and fast-moving data in a way that it cannot be examined with standard technology tools. In this context, distributed data collection is gaining increasing relevance. One of the core elements to enable distributed data collection is information needed to interpret the content and structure of the data, which is referred to as metadata. In general, metadata is as follows:

- information that describes data, i.e. ‘data about data’ or ‘information about data’
- the descriptive, administrative and structural data that defines data assets
- identification of the attributes, properties and tags that describe and classify information
- irreplaceable tool to manage the entire data lifecycle, processes and procedures
- one of the fastest growing sub-segments of enterprise data management.

Moreover, each distributed dataset is based on a harmonised metadata profile. The DCAT-AP specification, which was published from the European Commission, is worth mentioning in this regard [4]. This specification defines how metadata elements shall be described, especially when it comes to the following:

- data elements (description of a dataset in a minimal yet adequate way)
- wordings and semantics
- predefined categorisations
- data field names

- data value type
- data field lengths.

For more details, see Annex A that includes a table that overviews metadata elements required for a harmonised metadata structure (data fields and corresponding descriptions) [5].

The specifications need to be followed by all data owners to ensure machine readability via cloud-based services. Therefore, a metadata description for the transport sector needs to include at least following major items:

- metadata information (including date of creation of metadata, metadata language)
- content information (including name of dataset, description of dataset)
- temporal information (publication date and date of expiry)
- geographical information
- contact information (including data ownership)
- conditions for usage
- access information (including data format, data structure, access url)
- quality information (including update frequency, data collection methodology).

In summary, it is important, that metadata standards for transport are established and adopted by public and private sector data providers to enhance the usability of their data.

4.6. MOBILITY DATA MARKET PLACES ENABLING BIG DATA ACCESS

Market places on mobility data are currently being introduced all over the globe. Especially Europe seeks to bring this matter into a legislative form, where individual countries are obligated to share transport related datasets, referred to as 'National Access Points' (NAP), in uniformly defined data format. The required datasets, as well as, the required data formats are defined in individual European Commission (EC) regulations.

In addition, a 'Nominated Body' is needed for the verification of compliance of provided datasets with respect to quantity and quality that needs to be defined.

The European example shows that the setup of related infrastructure where data of different categories are made available for different stakeholders is a key enabler for new services. In many cases, such infrastructure is a kind of catalogue service, where data is described in detail and links to the data owners are provided. It is often the case that due to lack of knowledge of existing data, the setup of new services could be unnecessarily complicated.

Ideally, a data marketplace is set up in a way that metadata is also available in a machine-readable format and search engines are able to support filtering of relevant datasets.

Case Study: Germany on providing Data via a Mobility Data Marketplace (MDM)

In Germany, a single point of access to mobility data was set up called the Mobility Data Marketplace. The MDM brings together providers, users and refiners of traffic data. It is a neutral platform and it ensures transparent terms and safe technical standards, which makes simple to provide, search, and subscribe to traffic-related data.

When providers and users want to do business with each other they can exchange data via standardised interfaces and proven communications procedures. Both parties benefit from MDM's technical and organisational support, exploiting the full potential of the data.

Case Study: China on giving access to data of several stakeholders

A comprehensive traffic service big data platform was the specific result of Chinese transportation industry and internet enterprise cooperation. By using the cloud service technology of BAIDU, which is one of the biggest internet enterprises in China, a national data sharing and open platform was formed for public authorities and private organisations. In essence, it is a data exchange market where all the participants could get raw data, analysis results and applications they need based on valuable, high quality and affordable data. The data types available from the platform includes traffic flow, traffic incident, transportation operation and travel behaviour data for bus, taxi, and railways in urban areas, as well as, long-distance bus, airlines, hubs, and parking. Up to now, 10 provincial transportation departments, 11 internet companies and scientific research institutions have already joined the platform. The efficiency of transportation information resources application and transportation service level are improved with the help of data standards and operational regulations.

5. APPLICATIONS

Already today, several applications (services) are using big data for improving traffic management. This chapter will examine a number of example applications wherein big data is utilised to improve operations in various transport modes. While the major focus will be on road network operation, some examples on public transport as well as on emerging services in the area of sharing mobility were also included.

5.1. ROAD CASE STUDIES ON PROVEN AND EXISTING APPLICATIONS

5.1.1. Improved traffic management

Big data has the potential to improve traffic management. One key change when adopting big data in traffic management is having more data to calculate the traffic status and road incidents, which enables validation of the impact of implemented measures. This is expected to improve the quality of service (QoS) to end-users.

The other change brought about by big data is the ability to monitor not only single road stretches but the complete or a large proportion of the road network. This is particularly the case, as data from third parties can cover most of the primary roads, as well as, significant parts secondary road networks. Previously different parts of the road network tend to be managed by different entities operating in silos. With a more comprehensive visibility of road network performance, it is now possible to manage the network more strategically and the whole road network can be managed as an integral network resulting in more effective traffic management strategies.

Case Study Czech Republic on big data processing

In Czech Republic an innovative “dynamic mobility model” (called Rodos) was developed, which uses a range of data to predict traffic flows and to enable effective traffic management decisions. The Rodos Crisis Management (RCM) collects and analyses the following:

- anonymised user data about the distribution of mobile phones and their movements in the network
- floating car data (FCD) from company car fleet management databases (with 140,000 vehicles)
- all traffic detectors
- data on truck movements from the electronic tolling system

The above data are collected in near real-time. The data is overlaid with weather information, resulting in a fusion of data sources.

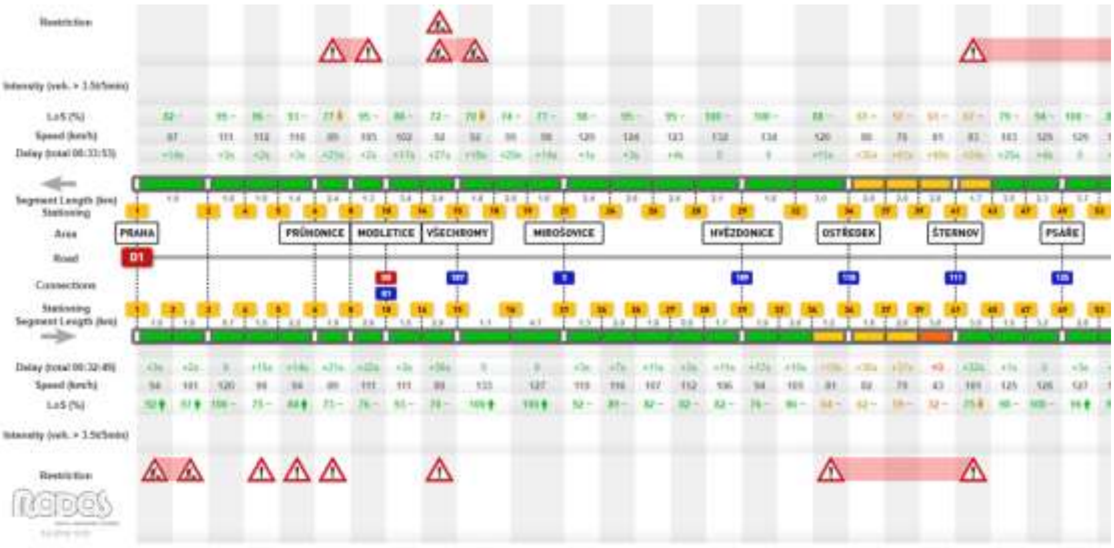


Figure 7: Level-of-service calculation based on big data analysis [6]

Output data of Rodos describes current traffic flow dynamics, including the following:

- average traffic flow and speed in each segment of the road network
- travel time for defined segments
- delay in identified segment (relative to free flow travel time)
- average free flow speed on segments
- weather-related traffic information (planned for latter stages of Rodos).

For more details, see Annex B.

5.1.2. Improved road network operation services

Collecting and processing significant amounts of data would help road network operators to validate the performance of current operations and further improve level-of-service. Big data will provide the inputs to better calibrate traffic models resulting in better tools for road network operations. It is therefore expected that big data processes will result in improved end-user services, as demonstrated in the following examples.

Case Study: United States (Utah) on improved traffic signal performance

In Utah, automated signal performance measurement was introduced. Signal performance metrics were introduced to show real-time and historical performance at signalised intersections. This allowed traffic engineers to measure what they previously could only model. Accurate decision-making about signal performance and timing helped signal management personnel identify vehicle and pedestrian detector malfunctions as well as monitor performance of timing plans. The 1,200 signals operated by Utah Department of Transportation traffic signals gather approximately 12 GB of data per day. Data processing is automated to generate the arterial performance measures.

For more details, see Annex C.

Case Study: Netherlands on reducing rush hour congestion

The Dutch approach on reducing rush-hour congestion is to encourage commuters not to drive during hours through a reward system. The reward may be in the form of cash, credit on public transport cards, loyalty points that can be used for online purchases, gaming elements, parking credits, and others. These are awarded to travellers who avoid driving during rush hour periods of a defined network or corridor.

The program has been tested since 2011 in Rotterdam (Netherlands) with 12,000 participants. The program awards a personal budget to every participant, which has been calibrated, based on their average number of trips during rush hours prior to enrolling into the program. Potential participants were identified, invited to the program and agree that their driving behaviour be monitored for the defined corridor or network. Identification and monitoring of travel behaviour of participants was conducted using a combination of technologies (i.e. ANPR cameras, GNSS smartphones, and on-board units) while ensuring participants privacy. 40% of the participants have effectively reduced their trips, resulting in a decrease in traffic by 5 to 10% during rush hours.

For more details, see Annex D.

Case Study: Australia (Sydney) on post-implementation analysis

The Pinch Point Program of the Roads and Maritime Services aims to reduce congestion at targeted traffic points, intersections or short lengths of road at which a traffic bottleneck exists. A methodology and tool to assess the impacts and effectiveness of pinch point projects was developed. The methodology utilised a combination of traffic count data from traffic signal detectors and travel speed data from probe vehicles to calculate performance indicators.

The assessment of the weekend clearway scheme on Victoria Road demonstrated that the weekend clearway scheme resulted in significant travel timesavings. The study also demonstrated that with emerging data sources such as probe vehicle data, it is possible to perform post-implementation analysis effectively and relatively easily.

For more details, see Annex E.

Case Study: London Orbital (UK) on the use of near real-time dashboards to improve operations performance on the M25 London orbital motorway

With increasing constraints, operators of infrastructure are looking for new means of improving efficiency and quality of service. Analysis of big data is a means of achieving both contractual compliance and improved performance by using information collected by operational systems. The application of big data was successfully implemented on the London orbital motorway in the UK, which included 440 km long highway section, including 220 km of the M25. On this highway, information from varied data sources was combined to generate real-time alerts and information dashboards.

All of the data that were collected through operational systems are centrally accessible via a 'data warehouse' which were then used to generate performance dashboards with automated early warnings of any significant changes observed in the operational performance data, such that it would allow managers to implement mitigation measure, when needed. Typical measures include incident response, defect management and winter maintenance.

These dashboards are accessible through the company intranet and allow users to select specific aspects of the information and to examine the data in more detail. This allows quick and easy access to information. These interactive dashboards can also be generated and transmitted automatically via email to specified recipients at set intervals.
For more details, see Annex F.

5.1.3. Automated weather detection

Closed Circuit Television (CCTV) systems are already part of existing camera-based traffic monitoring systems. CCTV systems are either used for observing stretches of road or to monitor specific sections of dangerous roads and bottlenecks. With the technological development, CCTV systems are becoming more intelligent with new functionalities such as traffic flow counting, incident detection and weather detection. These new generation of intelligent CCTVs would contribute to improved traffic management and safer roads.

CCTV have also an emerging role in weather-related detection services. It is however noted that inputs from CCTVs are only one of the many sources of data for weather detection. Additional meteorological data from third parties are additionally used for weather detection and weather forecasting for road user services.

Case Study: South Korea on improved weather monitoring

At the Pyeongchang (South Korea) Winter Olympic Games, which took place in January 2018, improved weather information systems were used. A combination of data coming from CCTVs and third parties were used to improve safety along the Korean Expressway network. Weather reports are provided by the Korean Meteorological Administration. Traditionally, these reports are prepared for regional areas. These regions include not only the heavily populated areas, which are most vulnerable to dangerous weather changes, but it also includes less populated areas.

To improve the format of the weather reports, digital weather reports were prepared wherein the road network links (i.e. linear units) form as the basis of weather information. This approach was a significant change in reporting format because it was the first time that linear features rather than areas/regions were the basis of weather reporting. To ensure that the reports are reliable, Automatic Weather Systems (i.e. meteorological stations) would have to be installed at 2-km intervals along roads. However, mass installation of AWS was not financially viable. Therefore other options were considered, particularly using data coming from other sources. It was therefore proposed to use CCTV systems that have already been installed on roads to monitor weather conditions. This proposal drastically reduced the costs.

Data from the Korean Meteorological Administration and data extracted from CCTVs formed the basis for weather monitoring along the national expressways. In addition, a real-time notification system was developed that alerts drivers of prevailing weather condition. Thus, the South Korean Expressway established a system that utilises existing surveillance cameras to observe and report on the weather.

Figure 8: South Korean Expressway weather reporting system configuration

5.1.4. Improving planning and operations tasks

Big data is not only used for services to travellers and network operators in real-time, but it is also used for evaluating the performance of the transport network including analysing the effectiveness of traffic management plans. In this context, big data has the potential to identify bottlenecks within a network and facilitate refining measures to improve network operations. Potential measures resulting from the big data analysis include construction of new roads, improvement or set-up of new ITS services such as dynamic speed limits, dynamic overtaking bans or adjustment of traffic signal plans.

Case Study Australia (Western Australia) on improved planning and operations tasks

A Network Performance Analysis for Perth was developed in 2015 and this process was repeated annually to provide updated information for Main Roads' congestion response strategy. The objective of this analysis was to characterise the performance of the Metropolitan Perth roads network and compare performance with the previous year. The findings of the analysis were utilised to develop Perth's congestion response strategy. It assisted the road agency to direct its efforts to the most congested sites leading to a targeted hence more efficient use of limited resources to address congestion hots spots. [15]

The analysis utilised traffic count data from traffic signal detectors and freeway data stations. This is supplemented by travel speed data from probe vehicles. Traffic volumes and travel speeds were compiled and were utilised to identify congestion patterns to inform the congestion response strategy for Perth.

The analysis examined roughly 58,000 of individual links (covering 4,453 km) of metropolitan road network and at hourly intervals on weekdays. Key performance indicators examined included various aspects of demand, travel speed, travel time, delay, reliability and cost of congestion.

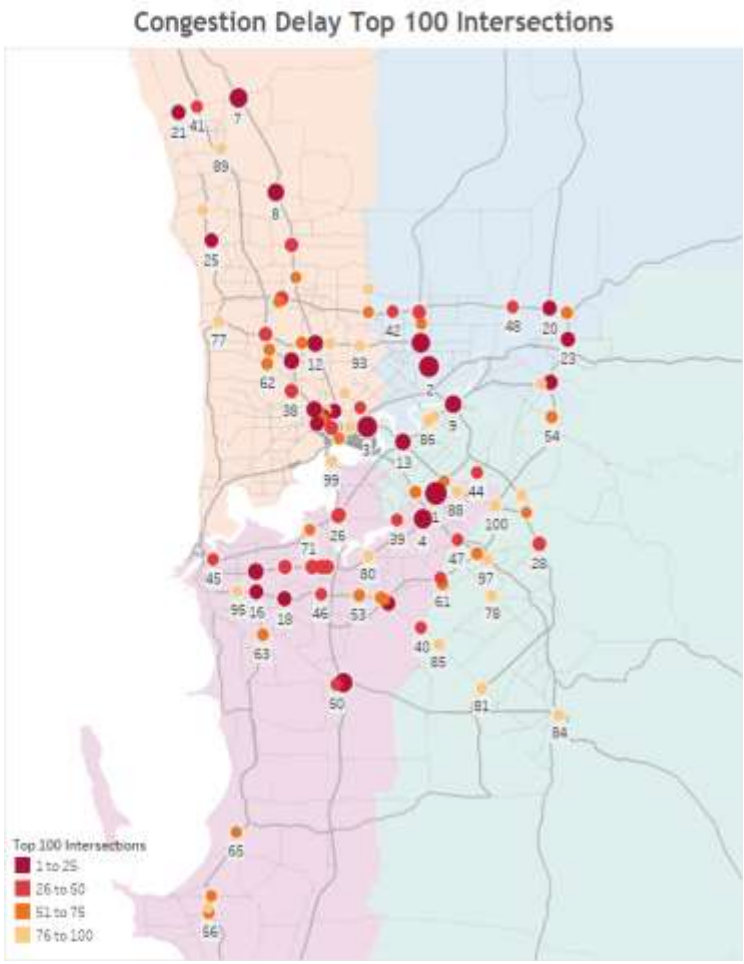


Figure 9: Identification of the most congested intersections in Perth for the development of a congestion response strategy [15]

5.2. ROAD CASE STUDIES ON EXPECTED FUTURE APPLICATIONS IN THE AREAS OF PLANNING; DESIGN; AND OPERATIONS

5.2.1. C-ITS as driver for improved road network operations

The connected vehicle is seen as a big data provider for road network operations in the near future. Through C-ITS services, it is expected that vehicles will not only receive data from road network operators on the current road status or safety critical issues, but the C-ITS equipped vehicles will as well provide data on the road status (traffic speed and traffic flow data) as well as on road events (including slippery road surface, weather conditions, or accidents). Hereby, different existing communication channels (long- and short range) are used for bidirectional data exchange. This mixture of communication channels is called "hybrid communication". In such a scenario, the connected vehicle will receive data, and in parallel extract data from in-vehicle systems (e.g. Controller Area Network – CAN-Bus) and transmit it to road network operators. Details can be found in the Technical Report provided by the PIARC Task Force B.1 Road Design and Infrastructure for Innovative Transport Solutions.

Case Study Europe on C-ITS deployment

In 2016, the European Commission has published the C-ITS strategy with the goal to initiate the deployment of so-called Day-1 services across Europe. In parallel, European countries have committed themselves to have a first road stretch prepared for C-ITS services, where data is exchanged between vehicles and road network operators. Additionally, car manufactures, especially the Volkswagen Group, have announced to start the deployment of C-ITS equipment in serial vehicles in Europe by 2019. Key European actors are committed to C-ITS deployments based on direct data exchange between road network operators and car manufactures.

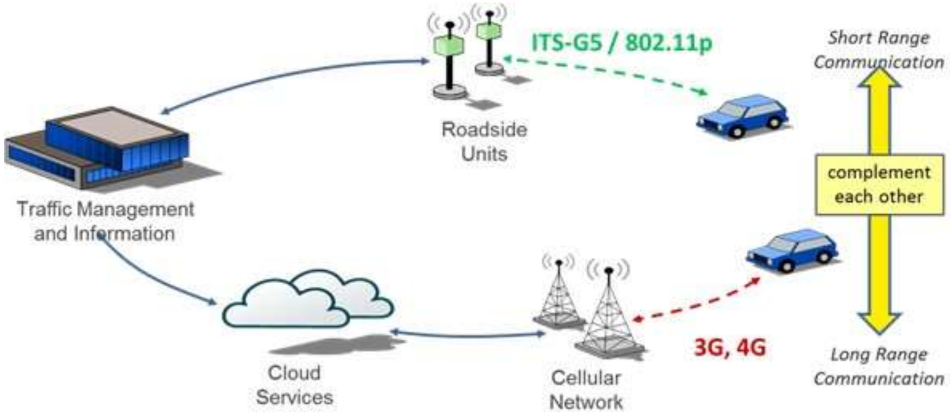


Figure 10: Hybrid communication for data exchange between road network operators and connected vehicles [7]

The basis for enabling such data exchanges in a harmonised and interoperable way is the common agreement on interfaces for data exchange. Hereby, the car industry is harmonising their efforts in the Car-to-Car Communication Consortium [8] and the European countries are working on deployment specifications within the C-Roads Platform [7]. In all related initiatives, it is of high importance to follow a legal framework especially when it comes to privacy issues.

In the end, C-ITS will demonstrate its benefit, not only by providing real-time safety critical services to the traveller, but by using vehicle generated data as additional source to road-side sensors for improved traffic management. By assuming that all vehicles along a specific road stretch will regularly communicate with C-ITS stations owned by road network operators, huge amounts of data will be generated, processed and analysed. This will help to fine-tune traffic management strategies and will result in improved services to road users.

Case Study Japan on Electronic Toll Collection

In the context of C-ITS solutions which is in essence communication between vehicles and the infrastructure, electronic toll collection (ETC) systems can also be included as part of C-OTS as they are also based on vehicle-infrastructure communication. In 2014 Japan introduced a new generation of in-vehicle on-board-units (OBU), called ETC2.0 OBU, that was developed as vehicle-to-infrastructure cooperative ITS. It uses the 5.8 GHz DSRC communication system that allows large capacity communication between vehicle and infrastructure at high speed. ETC2.0 OBU also enables the collection of in-vehicle data (also probe data called) on travel and behaviour information by adding GNSS functions to basic applications such as Electronic Toll Collection, safe driving information provision, and dynamic route guidance. As of August 2016, ETC2.0 equipped vehicles reached about 1 million, and the number is still growing.

The analysis of ETC2.0 probe data enables deriving information on road traffic with high degree of accuracy, and it clarifies current traffic situations and issues. Japan has deployed specific services based on big data for improving safety at congestion hot spots as well as at frequent sudden braking spots. Currently Japan is working on initiatives for “Smart Road Use” based on the analysis of ETC2.0 probe data. Additionally, Japan currently promotes research and development for automated driving based on big data collected via C-ITS services.

For more details, see Annex G.

5.2.2. The new role of Road Network Operators in a MaaS system

Mobility as a Service (MaaS) is one of the new paradigms based on big data management. Within a MaaS ecosystem, the role of a road network operator will slightly change. Managing the network will remain the core task for an RNO. However, services to end-users such services including traveller information, trip planning and in future maybe even tolling services might be integrated into a tailored MaaS end-user service as described in chapter 2.2.2.

MaaS, as a new way of organising and integrating transport, will require new types of and more extensive collaboration between transport service providers, public as well as private, and other stakeholders. Collaboration and cooperation between all actors in the field of mobility is a pre-condition to enable MaaS. Therefore, road network operators need to think on how to provide services and data to their travellers not only via direct links (e.g., via variable message signs) but also through MaaS operators. In such a changing ecosystem, new MaaS service providers will enter the market, both as unimodal providers (such as the recent emergence of transport network companies like Lyft and Uber) and multimodal providers or travel brokers (like UbiGo or MaaS Global).

In such a new environment, interfaces with the private sector need to be established to achieve the goals of local, regional or even national mobility goals. New mobility policies need to be established to develop a political vision and urban mobility objectives based on strategic alignment between all key public and private stakeholders of the extended mobility ecosystem. Such strategies need to be underlined by priorities as well as investments by the public sector.

5.3. PUBLIC TRANSPORT CASE STUDIES

This section highlights use of big data in several applications on public transport. Big data like transit (smart) card data, GNSS, cell-phone call records, online social media, smartphone apps data, and others are some of the widely discussed datasets in public transport analytics. Data generated by these sources are used for planning purposes as well as operation.

Smartphone apps can be useful in data collection for receiving direct feedback actively given by travellers, which is only possible if travellers volunteer. On the other hand, anonymised cell phone call-records are huge in volume and can be used to generate origin-destination (OD) matrices. However, the challenge stands for anonymised cell phone data to develop algorithms for filtering modes of travel. Considering the increasing smartphone usage and digital penetration, it is safe to infer that these big data sources will be promising and useful in reflecting gaps in transit services.

Data coming directly from transit agencies like smart transit card data and GNSS data from transit vehicles are much more credible and authentic reflection of transit in general. While this is a type of passive data, social media provide rather active data that can be used at evaluation stages for faster and direct feedback from users.

Table 4 – Data sources for public transport

	PLANNING LEVEL	OPERATIONAL LEVEL
Smart Card Data	1. Infer turnover rates, trip rates and the proportion of linked trips 2. Improving bus network planning by using smart card data as representative of passenger demand	For service quality monitoring on critical pathways, enhancing utility of current busway and selecting potential bus rapid transit (BRT) corridors
GNSS Data	Infer transportation mode use data from raw GNSS data	Assess more precisely the difference between the scheduled timetable and the actual movement of transit vehicles in the network
Online Social Media Data	1. Identify location and travel patterns 2. Evaluate transit rider satisfaction.	

5.3.1. Smart Card Data

Transit smart cards used for payment in public transport systems are used to create OD matrices and for evaluating transport systems. This has been done in several capitals, including Beijing, Santiago de Chile and in Istanbul for the Bus Rapid Transit (BRT) system. Singapore's Land Transport Authority used smart card date together with data from the telecom operators to analyse travel behaviour on the public transport system. The data from the telecom operators was used to supplement the first and last mile information to the public transport OD matrices [9].

Smart card data can be used for several purposes in transit planning. At the strategic level, smart car data can be used in the following areas with the aim to make public transport more attractive and competitive compared to other transport modes:

- Future network planning, including adjustments to the network topography
- Understanding user behaviour and improving user trust in public transport
- Adaptation of fare policy based on trip rates, the proportion of linked trips and turnover rates

At the tactical and operational level, including the planning of an effective public transport services, smart card data can support the following:

- schedule the coordination between different public transport modes (bus, tram, and metro)
- organise transfers more effectively
- adjust time tables
- getting better knowledge on alternative routes within the public transport network
- detect incidents quicker and more effectively

Smart card systems also ensure that transport service providers have access to larger volumes of personal travel data and knowledge of their most frequent customers. This enables them to link these personal data to the individual card and/or traveller giving them access to continuous trip data covering longer periods than it is possible to obtain using traditional transport data sources.

Case Study London (UK) on public transport network planning [10]

Smart card fare payment data can be of value in improving public transport network planning by being representative of passenger demand across the public transport network. The use of such data has been implemented in the city of London.

New information was derived from forming complete journeys using smart card data. The results can be divided into two categories:

- contextual knowledge about a route or station that may be quantitative or qualitative
- quantitative inputs to the cost-benefit models currently employed to evaluate frequency, capacity, or restructuring changes to bus routes.

Smart card data can be used to expand information on passenger demand, which might include passenger flows between intersecting routes to provide support for direct links that reduce the need for transfers. In addition, transfer volumes from bus routes to an underground station can be analysed to show which routes are the most important means of accessing the station and adjust station design and/or bus routing accordingly. The comparison of underground-to-bus transfer times can be used for optimisation of bus frequencies at an underground station, to highlight reliability or crowding problems. Evidence of multi-modal public transport journeys (e.g., bus-underground-bus) can support route redesign such as creating direct bus links that reduce the need to transfer and relieve congestion on the underground. Finally, the identification of repeated daily individual passenger travel on a route may help to indicate strong reliance on that service.

Three transfer combinations – bus-to-underground, underground-to-bus, and bus-to-bus – have been considered in order to develop recommendations for maximum elapsed time thresholds to identify transfers between stages of the journey. Raw data collected from the card readers on each bus (entry only) and at each underground gate (both entry and exit) are compiled into a sequenced journey table by Transport for London (TfL).

TfL used the smart card data to estimate network-wide origin-destination information for all travel in Greater London. This resulted in time thresholds for transfers in the London network across the whole network, as follows:

- underground-to-bus transfers are from 15 to 25 minutes between Underground station exit and bus boarding;
- bus-to-underground transfers, including bus in-vehicle time, are from 30 to 50 minutes between bus boarding and underground station entry; and
- bus-to-bus transfers, including bus in-vehicle time and waiting time for the second bus stage, are from 40 to 60 minutes from first bus boarding to second bus boarding.

Case Study Brisbane (Australia) on public transport network planning [11]

In Brisbane, smart card data has been used to analyse and compare BRT passenger travel behaviour with non-BRT bus trips. The objective of this analysis was to provide data for service quality monitoring on critical pathways, enhancing utility of current busway and selecting potential BRT corridors. Six months of smart card data were used. The smart cards provide information on name of operator, route ID, date, direction boarding and alighting stop and time, trip ID and smart card ID. This data was combined with general transit feed specification data, since the smart cards do not provide geographic information (e.g., the coordinates of bus stops).

The smart card data was first pre-processed to reconstruct bus trip trajectories and flow-matrices for BRT and non-BRT trips. Individual flow matrices for BRT and non-BRT trips between all the bus stops were created. This was done for workdays, school holidays, Saturdays, Sundays and public holidays. The results allowed the comparison of characteristics of BRT trips with non-BRT trips.

Case Studies South Korea on public transport network planning

In Korea, after the public transportation reform, which took place in 2004, the smart card-based fare collection mechanism was introduced in major metropolitan cities like Seoul, Busan, Incheon, Dae-Jung, Daegu, Kwangju metropolitan areas for bus rapid transit, regular bus and other rail services. This not only improved the convenience of users but lead to better decision making of operators of public transport modes for maintaining headways and improving the flexibility of services. This means that data accumulated is being widely used for public transport policy making encompassing operational scheme improvement, route planning, and subsidy setting decision.

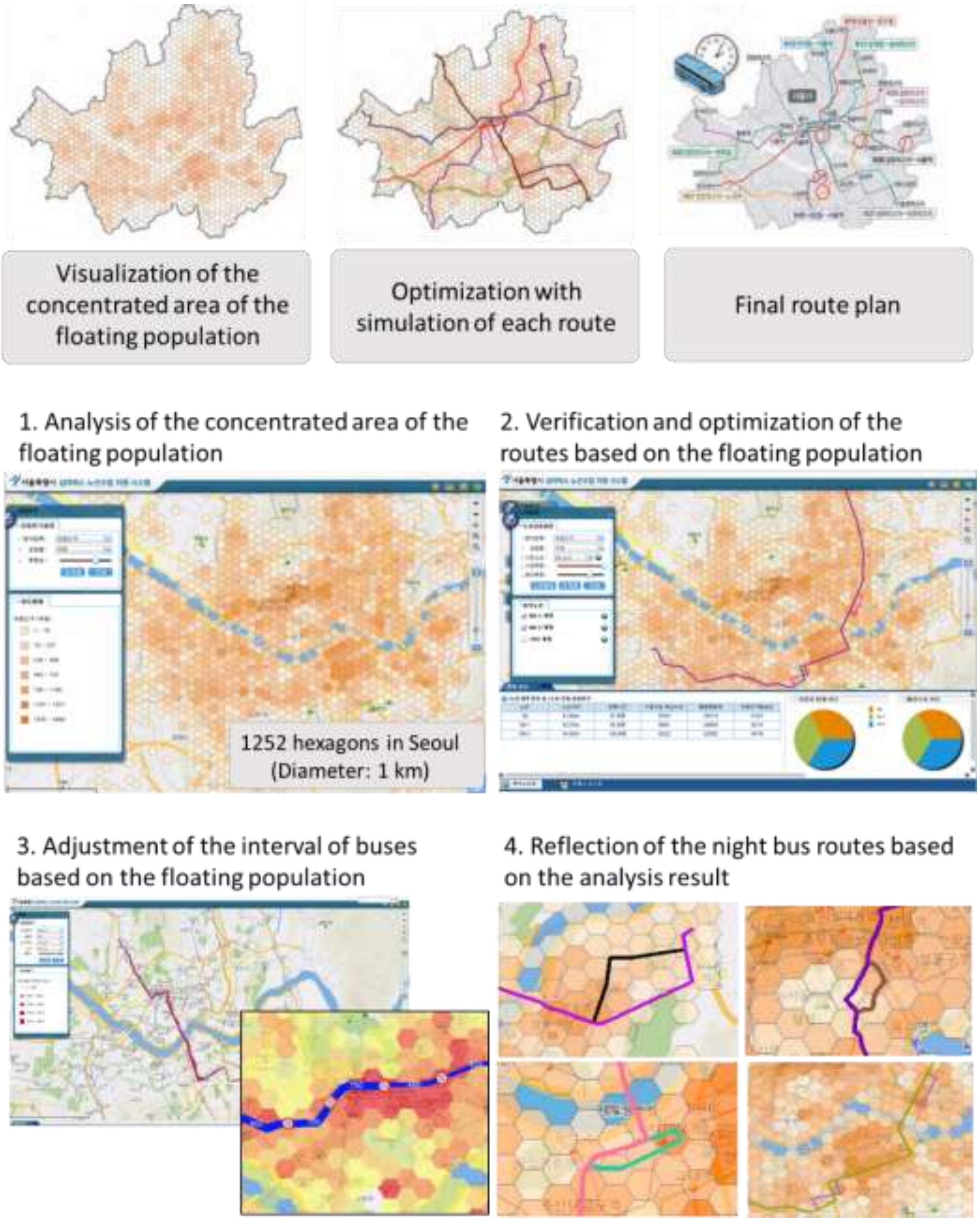


Figure 11: Process of determining routes, intervals and stations of the night bus in Seoul [12]

Better planning and operations were possible since each smart card transaction data contains each passenger's origin and destination information (time and location) to calculate fare based on distance travelled. To do so, users have to tag their smart cards while boarding and alighting to/from vehicle or station. This is being done nowadays in major Korean metropolitan cities.

Recently, smart card data from the Seoul metropolitan area was used to initiate express service of buses. An optimal limited-stop bus routes selection methodology was used for both reducing the line-haul time of bus corridor time and improving the competitiveness of public transit system in the area. More specifically, genetic algorithm was utilised to develop the model with different scenarios and the total travel time and operation cost savings from the implemented limited-stop service were computed. In case of the city of Suwon, travel time saving was estimated to be between 9 and 19 percent, if this method utilised.

The planning of on-demand services is another service that was set up with big data analysis. In the rural area, public transport in general is rarely provided due to low travel demand. Thereby the headway of any public transportation is rather long and sometimes services are very irregular. Considering such situation, instead of providing inadequate and wasteful supply of services, the Korean government will launch an on-demand service using existing taxi vehicles with subsidies. Data such as Bus Information System, Bus Management System, origin-destination travel demand from smart card data, and GNSS data containing administration border, road and transit network, major travel demand inducing facility can be utilised to establish a sustainable on-demand service plan.

For more details, see Annex H.

5.3.2. GNSS Data

GNSS data on the geo-location of transit vehicles can be used to analyse timetabling information. Such assessments are providing more precisely the difference between the scheduled timetable and the actual movement of transit vehicles in the network. In order to accomplish this, it measures the actual arrival times of transit vehicles at their stops and compare them against the corresponding times provided by the (theoretical) schedule. Even for lines of importance (such as buses joining top touristic hubs in the city), significant fluctuations between the actual position of transit vehicles and the original timetable can be evaluated. In addition, journey planners can become more accurate by the usage of GNSS data on the geo-location of transit vehicles.

GNSS data can be used to infer transport mode used. For example, if only GNSS data is available, analysing tools are available to automatically infer transportation mode from raw GNSS data. The transportation mode, such as walking, driving, etc., implied from a user's GNSS data can provide valuable knowledge to understand the user. It also enables context-aware computing based on user's present transportation mode. The detection approach is derived from following common sense knowledge of real world. People must stop and then go when changing their transportation modes. That means that there must be some GPS points whose velocities are close to zero during transition. In addition, walking is the usual transition between different transportation modes. In other words, the start point and end point of a walk segment could be identified as a change point with very high probability. Compared to uniform duration-based and uniform length-based segmentation methods, change point-based methods are achieving a higher degree of accuracy in predicting transportation modes.

Case Study Melbourne (Australia) on assessing the day-to-day variability of travel time [13]

A single bus route in Melbourne was examined with focus on day-to-day variability of travel time and its importance in determining service quality. The study analyses the nature and shape of travel time distributions for different departure time windows at different times of the day. Factors causing travel time variability of public transport were also explored using a linear regression analysis. The results showed that in narrower departure time windows, travel time distributions are best characterised by normal distributions. For wider departure time windows, peak-hour travel times also follow normal distributions, while off-peak travel times follow lognormal distributions. The factors contributing to the variability of travel times were found to be land use, route length, number of traffic signals, number of bus stops, and departure delay relative to the scheduled departure time. Travel time variability is higher in the morning peak and lower in the off-peak. The impact of rainfall on travel time variability was found to be significant in the morning peak.

5.3.3 Online Social Media Data

Online social media (e.g. Twitter messages) can contribute to transport planning. In cities like Singapore and New York, social media data is used to identify location data as well as travel patterns of individual travellers. In addition, social media data can be used to assess the acceptance of the transport system:

8. Do social media users describe transit planning, management, and services in a positive/accurate or negative manner?
9. Do differences in social media interactions influence the tone of the discussion surrounding agency services, planning, and public management on social media?
10. If so, should planners get more actively involved in social media to foster positive messages about public transit services and maintain civil dialog about patrons?

In all social media analysis the core element is the identification of relevant social media content about the specific transport mode from a huge amount of Twitter comments. Usually social media reflect more negative sentiments about public transit than do the comments about most other public services.

However, transit agencies may be able to influence the tone of those comments through the way they engage with social media. Transit agencies that respond directly to questions, concerns, and comments of other social media users, as opposed to merely “blasting” announcements, have more positive statements about all aspects of services and fewer slurs directed at patrons, independent of actual service quality. The interaction does not have to be customer oriented. Agencies using Twitter to chat with users about their experiences or new service also have statistically significantly more positive sentiments expressed about them on social media.

For social media analysis two related methods are used, i.e. machine learning and text mining. The text-mining method uses an algorithm to scan the Twitter text to match words found in the Twitter comments against an existing lexicon of positive and negative opinion words. Planners committed to a stronger role for public transit in developing sustainable and equitable cities have a stake in the social media strategy of public transit agencies. Planners should encourage interactive social media strategies. Even agencies that only tweet interactively a few times a day seem to have more civil discussions surrounding their agencies and announcements on Twitter than agencies that use their feed only to blast service announcements.

Social media analyses have some benefits measuring transit riders’ satisfaction compared to traditional surveys, including:

- The cost of data collection is minimal
- Data can be collected in real time
- User-specific needs can be assessed
- Data can provide meaningful insight as to why a particular sentiment is felt.

5.4. SHARING MOBILITY

Sharing Mobility follows the principle of sharing vehicles (e.g. cars or bikes). It can also involve sharing of rides, where car-drivers give other travellers the possibility of joining in for parts of a trip from origin to destination. In this respect, Sharing Mobility is a change of the paradigm from owning a vehicle via using a vehicle (vehicle sharing) towards using mobility (ride sharing). This change in paradigm is based on a networked society, and data is the key element to get access to shared services.

5.4.1. Vehicle-sharing

The most prominent example of shared services is with shared cars. Traditional rental car companies have been sharing cars. What is new, especially in urban areas, is the concept of sharing cars for single rides, which sometimes even last less than one hour. The idea is that users do not need to own a car any more, but that they use a shared car for rides, where public transport does not fulfil their needs. In urban context shared cars are seen by city authorities as an extension of the public transport services.

In the last years it turns out that such cars are not only available on fixed parking locations, but that they are used in a free-flow environment. That means that a car can be taken at any location, where it is parked and vice versa it can be left at any parking spot available.

To enable such services, cloud-based applications to the clients is a pre-condition. In addition, the ability to perform effective data management by the car-sharing operator is another pre-condition for such service. The car-sharing operator needs to monitor all cars, which are connected to a central cloud service by providing their location regularly through mobile communication networks. In detail following information is important to be processed by the sharing operator for enabling car-sharing services:

- Location of the cars to provide the right offer to the clients.
- Registered user management, only registered users will receive the allowance to use a car.
- Is the car booked?, it is important to know who is booking the car and in most cases a reservation (e.g. 15 minutes before the start of a ride) can provide “ride-security” for the traveller.
- Status of the car, it is important to ensure that the car is undamaged, tidy and refuelled.

The combination of the above data will allow car-sharing companies to give access to registered users in accordance to pricing models. The whole process of rent and return is usually done via mobile applications. So, all interactions are based on a cloud-based data and information transfer.

Ideally car sharing is not only driven by a single car sharing company, but it is embedded into a city’s mobility strategy where it is used as an additional service to make a city more traveller friendly and in parallel address environmental goals, such as reduction of CO_2 emissions by using electric vehicles for car sharing. In addition, big data processing allows monitoring and to analysis of rides done within a city, which helps to identify potential public transport routes. If specific routes are especially used by shared vehicles, this can be an indication of a gap in public transport service, which might be either a missing or inefficient. City authorities get inputs from big data analysis to improve their mobility plans for the city.

Like car sharing initiatives, the concept of bike sharing is becoming popular all over the world. Like car sharing services, bike-sharing works via cloud-based services to end-users. Similarly, big data processing is needed by bike-sharing companies. Further, bike-sharing services should be ideally embedded into a cities' mobility strategy.

5.4.2. Ride-sharing

In ride-sharing solutions, potential users contact a control centre to specify their destination, preferred time of travel, and any special needs. The centre then uses algorithms to identify the most appropriate vehicle operating that matches these requirements as close as possible. The vehicle could already be carrying passengers on compatible routes. It may be privately owned (such as a car) with the private owner simply giving a lift to the passenger; or, it may be a larger vehicle, perhaps a shared-ownership one. It may be a one-off or a regular journey. The dispatch may be carried out automatically or arranged through a website, perhaps involving an element of social networking [14].

In the past ridesharing was primarily offered by public transport operators as "on-demand-transport solutions" especially in loosely settled areas. In these situations, these services often were provided to a specific client group, for instance the elderly. Nowadays new service operators, like UBER or Lyft are offering ride-sharing services on a commercial basis, and they competing against traditional taxi services.

For ride-sharing services, cloud services and the processing of big data is essential to ensure tailored services to end-users. A link to pubic services and to a cities' mobility strategy would be very useful. Ride sharing services can be embedded into mobility platforms offering a bundle of services covering different transport modes and this could form the basis for new mobility solutions, like Mobility as a Service (MaaS).

5.5. Benefits in Planning, Design and Operation for Road Network Operation

Big data facilitates enhanced situational awareness, resulting in improved planning/design and operation. In comparison to traditional data, big data allows more up-to-date, comprehensive and disaggregated datasets with considerably higher value for planning/design and operation.

5.5.1. Benefits in Planning and Design

Planning and design of road network operation covers the following example tasks:

- Transport and traffic modelling
- Congestion hotspot analysis
- Transport service design (e.g., transit operation design)
- Post-implementation analysis
- Safe speed management
- Critical service planning (e.g., emergency service planning)
- Capacity and bottleneck analysis
- Freight management
- Toll design

To be effective, planning and design of road network operation require comprehensive and suitable datasets which are easily and cheaply accessed in a timely manner. Datasets relevant for planning and design are shown in Table 5. The table compares how these datasets are acquired traditionally and through big data.

Table 5 – Planning and design datasets for road network operation – traditional methods and big data

DATASET	TRADITIONAL DATA SOURCES	BIG DATA	BENEFITS OF BIG DATA	LIMITATIONS OF BIG DATA
Travel patterns, e.g., origin-destination	Traveller interviews	Device tracking, etc. smartphone applications and Bluetooth readers	Device tracking can offer higher sampling rates and lower cost than traveller interview methods	Device tracking requires suitable assumptions on trip purposes and to establish trip ends (e.g. no change in location for a certain duration is considered a trip end)
Travel speed and time	Travel time surveys	Network of permanent traffic counters (inductive loops), Bluetooth-sensors, Device tracking	Broader temporal and geographic coverage and lower cost	Big data methods are already widely adopted for travel speed and travel time data
Traffic count	Traffic count surveys and temporary traffic counters	Network of permanent traffic counters Device tracking in sections without permanent traffic counters	Broader temporal and geographic coverage Supplementary sample count data from devices can improve accuracy of counts for road sections without permanent traffic counters	Higher investment cost in installing traffic counters. However, typically traffic counters are installed as part of traffic control. There is a need to maintain traffic counter assets to maintain data quality and availability. Use of sampled tracked devices is still being developed, but if samples are high enough they could reduce the need for permanent traffic counters.
Spot speed	Spot speed surveys and traffic speed detection devices, such as traffic speed enforcement cameras	Device tracking	Wider coverage of device tracking and lower cost. Device tracking could avoid bias in data, as motorists tend to be aware of the location of permanent traffic speed enforcement cameras	Quality of spot speed data from device tracking may not yet be as accurate as purposely set-up spot speed survey (e.g., inductive loops)
Axle loads	Weigh-in motion stations	On-board mass monitoring	Wider coverage On-board mass monitoring may be applied as part of a road pricing and/or access management	On-board mass monitoring requires the installation of specialist devices on vehicles May only cover a particular type of vehicle, such as high productivity freight vehicles

Table 6 illustrates examples of benefits in the application of big data to planning and design in road network operations.

Table 6 – Benefits in planning and design from big data

APPLICATION	BENEFITS
Congestion monitoring	The monitoring of congestions along the road network and the analysis of the collected data can assist the road agency to identify proper measures, as it was performed in Perth, Australia (see chapter 4.1.5)
Post-implementation analysis	Data from vehicles equipped with tracking devices that automatically collect and transmit location, time, and travel speed, can be used to examine a congestion mitigation program of the road agency (i.e. Pinch-point program). Such an analysis performed in Sydney, Australia [16] demonstrated the congestion relief benefits of the program and it assisted the agency on assessing the value of the program for continued funding and it also highlighted sites where potential refinements for existing measures can be considered.
Crash risk and capacity bottleneck identification	Based on vehicle generated data crash risk and capacity bottleneck identification can be performed, as it was done in Japan (see chapter 4.2.1). Even if not embedded in the vehicle, on-board units used for other purposes (e.g. tolling) can be used for vehicle data collection. The analysis of vehicle-generated data can lead to the deployment of services to road sections that were identified as potential crash risks and capacity bottlenecks based on e.g. vehicle braking data.

The application of big data for planning and design for road network operation is not necessarily based on reducing the cost of data collection. In some cases, the cost of data collection may increase. The benefits of big data are primarily on the quality and scope of data, such that planners and designers are making evidenced-based decisions that lead to better plans and designs. The value that big data could potentially offer is significant as a result of improvement in services to road users and improved efficiency in the delivery of these services. Still, the investment that is required is often relatively small, especially if the collection and analysis of data has been designed into existing services, e.g., compilation of smart card data.

5.5.2. Benefits in Operation

Operation in road network operations include:

- Traffic congestion monitoring
- Traffic control (intersection control, ramp meter, variable speed management)
- Incident detection and response
- Traveller information
- Enforcement
- Audit of tolls

Operation in road network operations require accurate data at frequent refresh rates to allow to maximise the effectivity of operational strategies and control. Datasets relevant to operation are shown in Table 7, which also compares traditional and big data sources.

Table 7 – Operation datasets for road network operation – traditional methods and big data

DATASET	TRADITIONAL DATA SOURCES	BIG DATA	BENEFITS OF BIG DATA	LIMITATIONS OF BIG DATA
Speed, volume and occupancy (SVO) Queue and/or density	Detectors CCTV	Device tracking	Device tracking allows more comprehensive coverage	Big data methods are not yet suitable due to low latency and low sample rates. Moreover, locational accuracy is not yet developed well enough for lane level SVO data
Real-time congestion and travel time monitoring	Modelling based on detector inputs CCTV	Device tracking	Device tracking provides a relatively accurate and broad coverage of congestion, especially for high volume highways (high likelihood of availability of a traced device in the traffic stream)	It is possible that there is no tracked device in the traffic stream to detect congestion, such as in local and collector roads. Data fusion on modelling based on detector inputs and device tracking may be suitable at the moment, until the number of tracked devices increase.
Incidents	CCTV cameras and phone reports SVO-based automatic incident detection methods	Device tracking to detect unusual congestion patterns	Device tracking coverage can cost effectively detect incident based on its effect on travel speed/congestion patterns	Same as above, device tracking penetration rates are still low to be a reliable source of congestion patterns for incident detection. At present, device tracking data may be used as supplement to traditional data sources.
Parking	Parking surveys	Parking sensors	Parking availability can be known in real-time and utilised for parking guidance systems	Parking sensors require funding and its installation would not be viable or practical in cases where parking availability is not an issue

Table 8 illustrates examples of benefits in the application of big data to operation of the road network.

Table 8 – Benefits in real-time or near real-time operation from big data

APPLICATION	BENEFITS
Unusual congestion monitoring	Sensors like Bluetooth readers can be installed throughout the road network to monitor congestion levels in real-time. Collected data is used to identify unusual congestion, which can point the traffic management centre personnel to review sites for potential issues. This is often verified through CCTV and data from the traffic signal control system. Such systems facilitate faster response to the mitigation of congestion, as it was e.g. don in Melbourne, Australia [17]
Motorway traveller information and incident detection	Sensors along the road network can monitor travel time in real-time. The information is used to inform motorists of traffic conditions to facilitate efficient route selection. The travel time data can also be used to flag possible incidents on the road network. Early detection of incidents allows faster response of incident response teams and that roadway capacity can be restored quicker, as it is done in Munich, Germany [18]
Traffic signal settings improvement	Data from signal sites can be utilised to show real-time and historical performance of traffic signals. The data allows traffic engineers to monitor and refine traffic signal settings as well as identify vehicle and pedestrian detector faults. This results in improved arterial performance, as it is done in Utah, USA (see chapter 4.1.3)
Operational performance improvement	Based on data collected through operational systems across the network, near real time performance dashboards provide automated early warnings of any significant changes in operational performance data. These warnings allow managers to put in place the required improvement actions and quickly improve performance in a number of areas such as: incident response, defect management and winter maintenance (see chapter 4.1.3)

Real-time application of big data for road network operations is still largely under developed due to low latency and low penetration rates. Research is currently being ongoing on real-time applications of big data and there are promising developments such as in incident detection, area-wide traffic control, predictive traffic management schemes and others. On the other hand, near real-time application of big data for road network is already finding use in traveller information and parking guidance systems. Big data application in operation of the road network would be able to unlock spare capacity in the network and allow for fast response traffic interventions to improve road network performance. As big data for road network operations improves in terms of quality and scale traffic management would transform from traditionally based on historical averages to one that is able to adjust to real-time variation in traffic conditions. It is also likely that traffic management incorporates aspects of predictive control strategies.

6. ENABLING ELEMENTS

A key factor for efficient planning, design, and operation of transport networks is the availability of relevant information at the right time for specific decision-making processes. Big data usage is one of the most effective tools to meet these needs of traffic managers and other stakeholders. There is no doubt that the data revolution will make a significant contribution to efficient traffic planning and traffic management, but there are related risks that need to be tackled and addressed. This chapter looks at the possibilities and risks associated with big data. The section identifies traditional and new risks and considerations for success, including skills, standards, governance, management, architecture, data quality, data ownership, security and privacy.

Research on public attitudes on the use of personal data has identified privacy, security and ownership as key concerns. Good governance and management of big data can improve data quality and reliability of results; ensure legal requirements regarding data privacy are met; and encourage public interest and involvement with, big data projects. Finally, standardisation in big data is needed for the long-term sustainability of designed systems and processes.

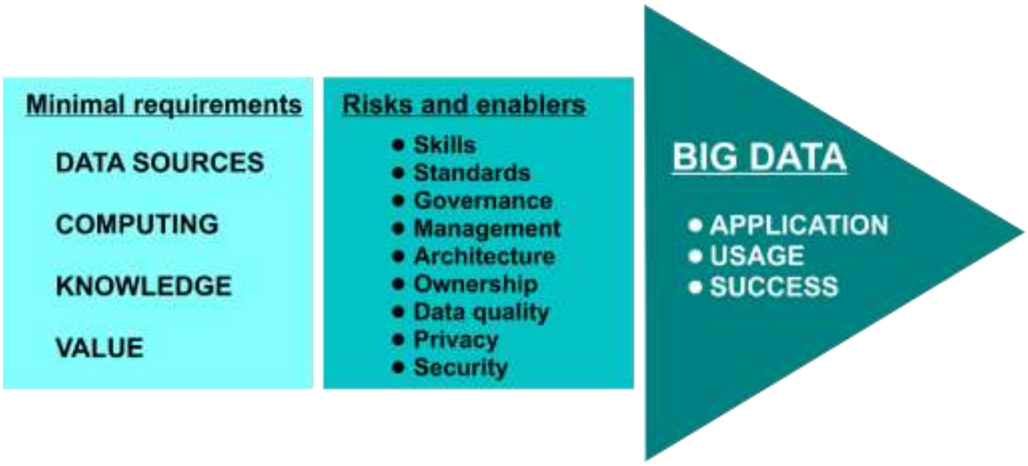


Figure 12: Enabling Elements

6.1. MAKING ORGANISATIONS READY (SKILLS NEEDED)

Data science aims to help an organisation turn data into useful and valuable information to support decision-making. With the development of big data, new forms of roles have appeared recently such as “data scientists” and “data engineers”. These roles and functions were formerly performed by “data analyst” and “business intelligence developers” who are responsible for platform development (e.g., data warehouse) to support the analysis of data. Data scientists and data engineers are part of a larger organisational team composed by managers, IT professionals and front-line employees.

6.1.1. Main skills in data science

The entire data life cycle concerning big data requires different skills, knowledge and expertise. The data life cycle is the set of processes in an application that transforms raw data into actionable knowledge. The classification of skills needed in data science is listed in the following and interactions are presented in Figure 13.

- Main skills in data science:domain expertise
- analysis & scientific method
- mathematics & statistics
- machine learning
- analytical applications
- algorithms
- data engineering
- advanced computing
- visualisation
- hacker mind-set

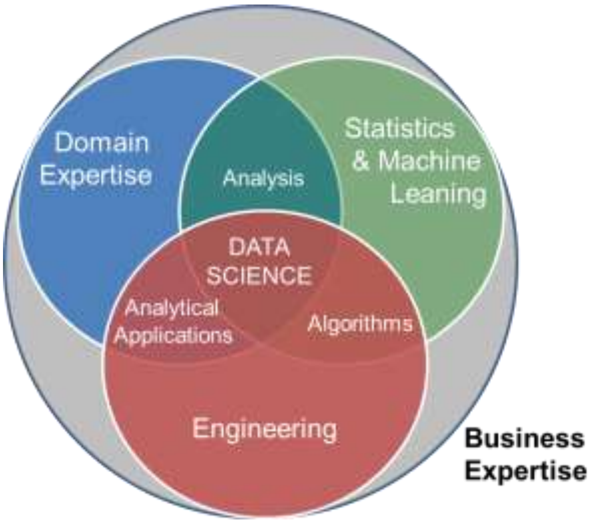


Figure 13: Skills Needed in Data Science [19]

Data science is an empirical science, which applies the scientific process directly on data, notably driven by a business need and to discover and to better understand the behaviour and parameters of a problem. Data science includes principles, techniques, and methods from many disciplines including data cleaning, data management, analytics, visualisation, data interpretation, and now including big data engineering.

Analytics is the heart of big data processing and its concerns with the discovery of meaningful patterns in data. It is one of critical steps in the data life cycle to produce valuable information for the organisation. The analytic process includes four fundamental steps (or questions) in order to lead to decisions and actions:

- descriptive (what happened?)
- diagnostic (why did it happened?)
- predictive (what will happen?), and
- prescriptive (what should be done?)

This is illustrated in the following figure. The operations or process in the figure may be related to either human or machine capabilities.

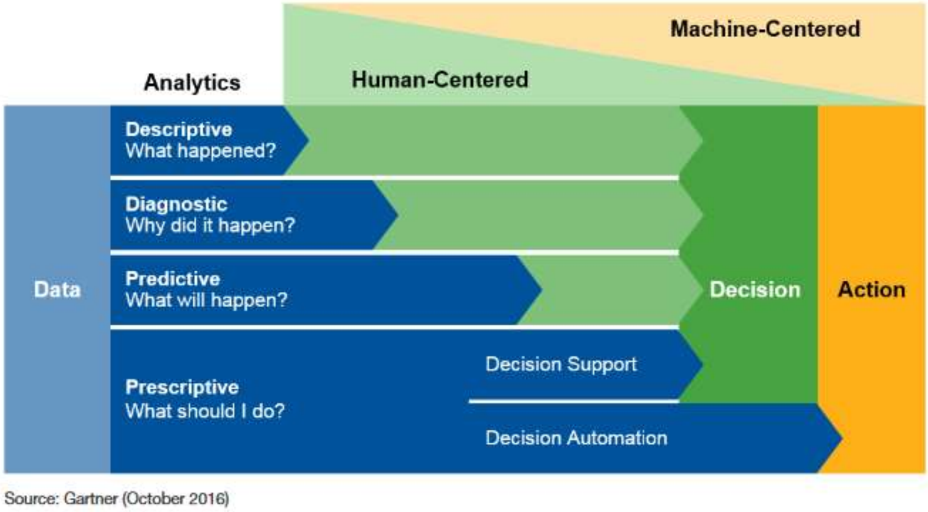


Figure 14: Four Analytic Capabilities [20]

6.1.2. Data scientist

Data scientists dedicate efforts and knowledge to the analysis of data to solve business problems. Their role is to transform data into valuable information and knowledge to support decision-making concerning business strategies to be adopted. The data scientist has the ability to put together this knowledge to produce valuable information in the transportation domain for the organisation.

A data scientist is a practitioner who has knowledge in the overlapping regimes of business needs, domain knowledge, analytical skills, and software and systems engineering to manage the end-to-end data processes in the data life cycle.

The data scientist explore data to analyse scenarios based on hypotheses by applying various methods and techniques in analytics, statistics, and machine learning to identify and solve critical business problems and prepare the results for interpretation by practitioners. This work includes the development of statistical models; discovering pattern in the data; and being able to relate various variables and datasets to interpret and explain findings using visualisation techniques.

The data scientist is a key expert human resource. This person must have a variety of skills in the transportation business domain, analysis, statistic and machine learning, analytical application, and in computer science (algorithms design, large data engineering). Some skill in geomatics is needed as well because localisation data is fundamental in transportation analyses and representation. Geomatics data is a key element to relate the various aspects of transportation such as population, weather, social welfare, etc.

Data engineer (or data analyst and architect in big data)

The distinction between the roles of data scientists and a data engineer is important. The data engineer's role is to find, organise, clean, sort and move data, and enable data scientists to leverage their skills to solve business problems. In this respect, a data engineer is the data professional who has main responsibilities to prepare the big data infrastructure. The data engineer designs, builds and manages the information infrastructure of big data. Their role is to provide data in a ready-to-use form for data scientists to execute queries and run algorithms on data for predictive analytics, machine learning and data mining.

Data engineers develop the architecture to analyse and process data according to the needs of the organisation and to ensure the effective performance of the system. This task includes the creation of data feed mechanisms to gather information from different source systems and complex queries to prepare data (consolidate, clean, and structure data) for usage by data scientists. Data engineers often work in analysis teams and are focused on design and architecture. Therefore, they do not need to know machine learning or big data analytics.

Other skills needed

The entire process in the management and exploitation of big data needs other complementary expertise to manage, support, and operate big data mainly in the field of information technology (IT), notably in software, hardware, and network. In the interaction between data scientists and data engineers, data science teams are fundamental to solve complex data problems. The skills in communication, presentation, investigation and interpretation are also very important given the complexity of interactions in big data systems.

Other skills required for the development, operation and management of big data are mainly typical resources that involve in information technology domain, notably in system development and information management. These resources are as follows:

- data manager
- security manager
- business architects
- data change agents

Specific skills needed within the transportation domain

Domain expertise is acquired and developed by transportation experts to understand technical issues related to transport and to identify potential solutions. This includes theoretical and practical aspects of transportation and is acquired through education and experience. To take advantage of the emerging big data field, it is also desirable for transport domain expertise to include big data and artificial intelligence.

The domain expertise in transportation network operations and management includes the following:

- transport network infrastructure planning and operation
- transport network service development (supply)
- transport network demand forecasting and usage patterns (demand)
- logistics planning in freight transportation
- social and economic aspects of transportation analytics via big data processing
- other aspects influencing mobility

6.2. STANDARDISATION

This section discusses standards related to big data. It also considers standards in the field of intelligent transport that have been defined for road transport operations that are useful for big data management and analytics. These standards will better define big data for users and facilitate the development of information for the management and operation of the road network. These standards include aspects of data such as meaning, structure, operation, organisation and management of a big data with consideration of security.

Standardisation consists of programs to minimise, simplify, and rationalise whenever possible which, if left alone, would become divergent, complex, or chaotic. The original aim of standardisation in the industrial field is to ensure the compatibility of products and provide an environment wherein customers are not limited to purchasing from a particular supplier.

Written rules defined by standardisation are generally referred to as standards. Typically, a standard has no binding power, as would a legal requirement, which means that standards are optional. Ordinarily, the standards should be defined based on an agreement among concerned parties.

Government agencies often mandate compliance with specific standards (mandatory standards) for the purpose of public benefit. The key roles of standardisation are as follows:

- securing the compatibility of products and assuring interfaces
- improvement of production efficiency
- assurance of quality
- accurate communication, promotion of mutual understanding
- prevalence of technologies from research and development
- assurance of safety and security
- reduction of environmental burden
- enhancement of industrial competitive strength, preparation of competitive environment
- promotion of trade, and more

Two families of standards are involved in the implementation and operation of big data to support road network operation as well as to link with the data that can be sourced from ITS, and they are as follows:

- Standards specific to big data: ISO/IEC JTC1 WG9
- Standards on the intelligent transportation system: ISO/TC204

ISO/TC204 covers the standardisation of data contents and formats that might apply to big data and applications using big data [21]. There is interaction and distinction between the ISO Groups above. Big data standardisation is mainly focused on the needs of the data providers (performance data, operations data, planning and modelling data, geospatial data, asset data, social data and weather data). On the other hand, ITS standards covers the outputs of information services supporting vehicle owners, data protection authorities, transport authorities, transport planners, transport service providers, and transport users.

6.3. PROCESS MODEL (BIG DATA REFERENCE ARCHITECTURE)

The following figure illustrates a big data reference model, which shows the various components, actors, and roles which are described in more detail in the following sub-sections.

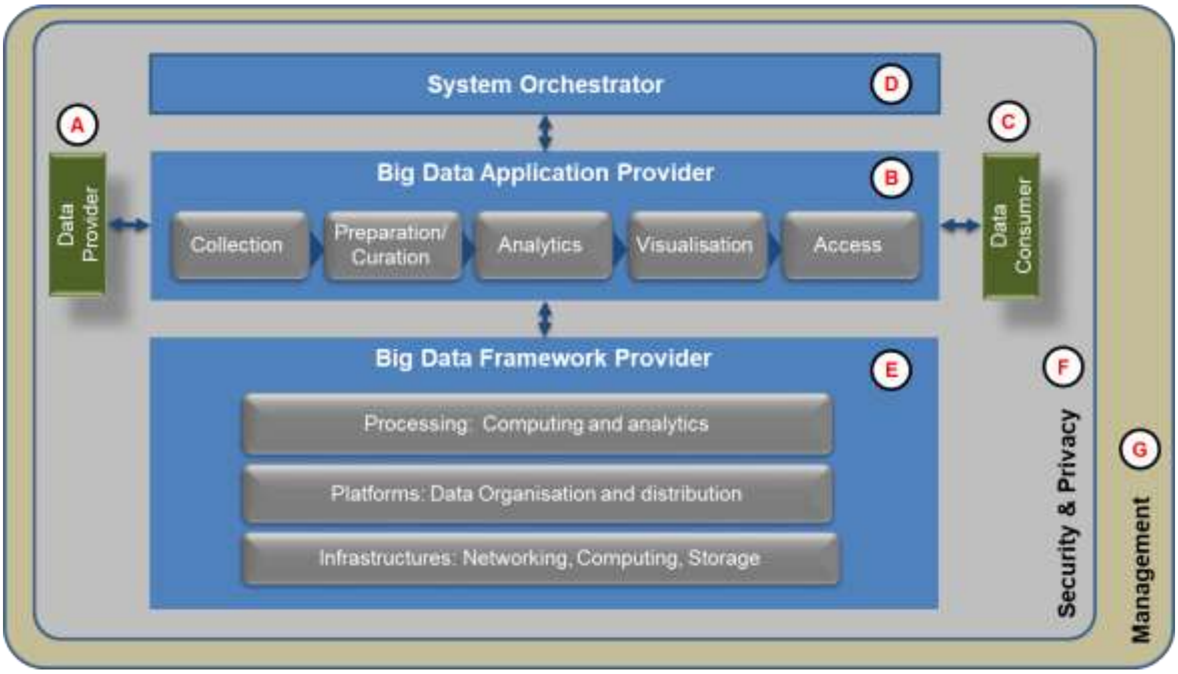


Figure 15: The big data reference model for the transport domain [recreated Figure from 22]

In big data, different roles are played by different actors depending on the development, operation, orchestration, and management of big data. Theses system actors can represent individuals, organisations, software tools, or hardware components.

In the past, data systems were typically hosted, developed, and deployed with the resources of a single organisation. With the advent of big data, the actors can also come from multiple organisations and the roles can be distributed accordingly amongst them.

Figure 15 identifies the basic elements of a big data reference model, which can be applied in the transportation domain. This model is a simplified version of the NIST (US National Institute of Standards & Technology) big data reference model [22]. The NIST reference model is used as well by the International Standards Organisation (ISO). The model is composed of five functional components and two frames. The functional components are connected by interoperability interfaces. The components that are referred to as "provider" indicate that they provide or implement a specific technical function within the system.

Table 9 - Components of the big data reference model

FUNCTIONAL COMPONENTS	FRAMES ENCOMPASSING THE FIVE FUNCTIONAL COMPONENTS
A) data provider B) big data application provider C) data consumer D) big data framework provider E) system orchestrator	F) security and privacy G) management

A) Data provider

The data provider makes data available for usage to the big data process or to other systems and users. This actor can be either part of the big data system, or can provide input from another system, internal or external to the organisation orchestrating the system. Once the data is within the big data local system, requests to retrieve the needed data will be made by the big data application provider and routed to the big data framework. The data may be collected from own or 3rd party sources, whether it is an online source (including browsers, sensors, detection devices, mobile devices) or an offline source (public or internal records).

The data provider's activities concern the data production process (data capture from sources, data persistence, data scrubbing, data annotation, and metadata creation) and other complementary aspects like access rights management, access policy contracts, data distribution application programming interfaces, capabilities hosting and data availability publication.

B) Big data application provider

The big data application provider executes the operation of the data life cycle based on the established requirements to meet security and privacy requirements. The general capabilities within the big data framework are then combined to produce "specific data" for the data consumer. In this respect the big data application provider must fulfil the following functions (refer as well to chapter 2.2):

- Collection: establish the data ingestion mechanisms to capture (obtain, load) data from the data provider and include exchange protocols and security, data formats, and metadata.
- Preparation: includes validation, cleaning, outlier removal, conversion, calculation, indexing, aggregation and summarisation, storage preparation, etc.
- Analytics: relates to rapid analytics, quick correlation or trend analysis to identify potentially promising hypotheses.
- Visualisation: which is distinct from the analysis, supports exploratory visualisation for data understanding (browsing), comprehending analytical results (real-time presentation of analytics and interpreting analytical results) and getting an overview of the situation (reports, summarisation)
- Access: to system data for the data consumer, including data export tools, consumer analytics code hosting and analytics as a service hosting

C) Data consumer

The data consumer receives the value output of the big data system. In many aspects, the data consumer has the same functionality as that the data provider brings to the big data application provider. After the system adds value to the original data sources, the big data application provider then offers that same functionality to the data consumer. The main functions of a data consumer are search and retrieve, download (export data), analyse locally, reporting, and visualisation of data output from big data analytics.

D) Big data framework provider

The big data framework provider offers the general IT resources or services required by the big data application provider to create and operate applications. The role of the big data framework provider has the most significant influence on the introduction of big data. The big data framework provider has to define the following:

- Infrastructure frameworks that concerns networking, computing, storage and environmental installation,
- Data platform frameworks that invloves data organisation and distribution,
- Processing frameworks that covers processing services, metadata descriptions, queries, temporal treatments, applications, messaging, and resource management.

Many applications for big data are based on hybrid technologies. Here several new support services need to be offered by the big data framework provider. These cover internal services and external services like cloud-based services (Saas – system as a service). New methods of data distribution and orchestration of data processes on parallel resources have been developed and introduced and need to be supported by the big data framework provider.

E) System orchestrator

A system orchestrator is responsible to provide the organisational requirements to the big data system, including rules, governance, architecture, resources, and business requirements, as well as monitoring or auditing activities to ensure that the system meets these requirements. The system requirements definition includes high-level design, data system monitoring, and the integration of the data application activities into an integrated operational system. The system orchestrator has a critical role in the management of big data by setting up the whole processes and defining the framework of the system.

The system orchestrator activities applies to several dimensions of big data management including business (goals, targeted actions, contracts and service level agreements, negotiation, cost analysis), governance (policy compliance, change processes, data stewardship and ownership), data science (functional performance of the analytics, data and analytics design), and system architecture (data process, software, hardware, logical data modelling and partitioning, data import/export, scaling).

F) Management framework

The big data characteristics of volume, velocity, variety, veracity, and variability demand a versatile management platform for storing, processing, and managing complex data. Management of big data systems should handle both, system and data related aspects. The management framework encompasses two groups of activities: system management and big data life cycle management:

- The system management includes: provision, configuration, package management, software management, backup management, capability management, resources management, and performance management.
- The big data life cycle management involves activities surrounding the data life cycle including collection, preparation/curation, analytics, visualisation, and access.

G) Security and privacy framework

Security and privacy considerations are a fundamental issue and affect all other components of the big data model. The security and privacy framework include all elements as all entities need to agree on the security protocols and mechanisms required.

H) Other functionalities

In addition to the functionalities mentioned above, some aspects deserve attention in the context of the big data, particularly in the case of diverse sets of data and the distribution of services and treatment on several technological platforms. These roles concern data governance and infrastructure management, which are as well in the scope of the management framework as well as the security and privacy framework.

- Data governance

Data governance is a fundamental element in the management of data and data systems. Its purpose is to administrate, formalise, and discipline the behaviour in the management of data. The definition of data governance covers the complete data life cycle and its different states (storage, processing, transaction, etc.). Data governance also considers the issues of privacy and security of persons and organisations. Data governance remains important no matter the size or complexity of the big data in terms of data volume, velocity, variety, veracity, and variability. Data governance will need to remain active throughout the data lifecycle so that it stays relevant to its stakeholders.

- Infrastructure management

Infrastructure management is related to hardware operation and maintenance, including security and privacy considerations. In any condition, infrastructure management must ensure resilience, redundancy, and recovery of the big data system. The infrastructure management includes threat and vulnerability management, mitigation qualitatively or quantitatively, configuration management, logging, malware surveillance and remediation, and network boundary control.

6.4. Quality criteria

Big data in transportation is used to measure a set key performance indicators (KPIs) such as travel time, number of crashes, CO_2 emissions etc. Big data tools are suitable for such measurements, but quality data is needed to ensure quality outputs. This section focuses on the basics of measuring and setting the quality criteria for data collection, preparation, analytics, and storage processes. It is essential to know and understand the quality of the data available. Big data processing and aggregation needs to apply a data quality assessment, otherwise outputs will be inaccurate and users' decisions based on the data risk that their needs will not be met and data providers' reputation will suffer accordingly. Transport network and service operators need to manage data to ensure it is available, reliable, accurate, and true. This means that relevant data standards should be established and input data needs to be monitored, controlled, and refined [23].

Big data is still a new concept and a uniform definition of its data quality and quality criteria has not yet been setup. Data quality depends not only on features of the data, but also on the environment where data is collected and evaluated including the processes, applications and end users. Only data that corresponds to the relevant use case and meets requirements can be considered as good and valuable data. Typically, data quality standards is initiated from the perspective of data producers. In the recent past (and nowadays), data consumers are both direct and indirect data producers, hence they were able to ensure that the data is of appropriate quality. With big data, there is a huge amount of data sources, hence data consumers are not necessarily the sole producers of data. It is therefore difficult to measure and ensure data quality. At any rate, it is important to ascertain that a big data system is usable based on how its data quality.

Different quality indicators may be developed depending of the phase of the big data process (i.e. data collection, data preparation, data analytics, visualisation and data access, described in chapter 1.2). For example, the complexity of the file structure where data is stored can affect potentially its readability, the integration with other data, and the level of data that can be disseminated which will be evaluated respectively at the input, processing and output phases (using different quality indicators). Indicative values of quality thresholds and calculation methods shall be set, including the following quality criteria:

- Availability
- Accessibility – access interface, open data (purchase data)
- Timelines – timestamps within a given time, arrival time, updated time, etc.
- Usability
- Credibility – correctness of datasets and values are within an acceptable confidence interval
- Reliability
- Accuracy – data is accurate and reflects the truth
- Consistency – repeated processing of the data will result in same or similar results
- Integrity – format is clear and structure of data is consistent
- Completeness – the scope of the data is adequate to target components
- Presentation quality
- Readability – content and format are understandable and clear.

6.5. Data ownership

Each data and data set always has its owner. Before using any data, it is always necessary to have the owner's permission. Much of the value of big data comes from combining data from different sources. Combining data in this manner can provide broader context. Transparency of ownership enables trust and control for data owners, as well as openness and utility for enterprises and society. Maintaining data provenance allows traceability through the data lifecycle and tracks data ownership and change of ownership.

Data privacy and the ownership of data are potentially controversial and multifaceted issue. However, if data is collected for creating value for users or society, people will be more willing to share their information.

In case of situations where large data systems are deployed, data ownership issue needs to be given careful attention. The data ownership problem can be so complex that it would go beyond the scope of this study, for this reason below are some basic recommendations that should be taken into account in the initial design of a system that will use data from different sources:

- The core solution is comprised of a "vendor neutral hub" — i.e. a platform that works independently of vendors and data owners where data can be captured, safely and durably stored, processed and distributed and finally shared, if permitted by the data owner.
- Data can be of great value hence companies should consider ownership and privacy issues before using big data results.
- In databases, ownership should confers the privileges to create, read, update, and delete all data, even whole datasets.
- Ownership should be an attribute of data (which may or may not be visible to users) that associates data with one or more entities that own or can influence what can be done with data.
- Data ownership must be addressed in the context of security and privacy of big data.

6.6. Privacy

The acquisition, collection, analytics and distribution of personal data including personal information play an increasingly important role in social and economic activities, innovation and economic growth. As the permeation of IoT and AI is expected to result in even more data acquisition and analysis in the future, the importance of arranging the framework for utilisation of personal data and information is important.

With vehicle data, it is evident that intelligent devices on modern vehicles already contain data and more advanced connected vehicle systems will provide even more data both on-board and outside the vehicle, or even on other vehicles. A more intelligent and communicating vehicle will generate and store large quantities of data some of which may be of a private nature. Some data might be of interest to the authorities, police or insurance companies. It is clear from international practices that as a guiding principle for building security and data protection in the development of software life cycle that it is based on the principles of "privacy by default" and "privacy by design".

Proper consideration of privacy issues must be based on reaching a consensus on the acceptable balance between loss of privacy and the achievement of objectives, such as seamless journeys. For example, using origin-destination matrices are probably the most problematic use case because the most valuable insights are drawn from analysing highly disaggregated data from individual travellers.

There are a number of definitions of privacy from that offered by Google to Article 12 of the UN Declaration on Human Rights (1948) [24]: "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."

6.6.1. International examples of personal data regulation

OECD

An international standard for personal data protection in probe systems had been discussed in the Technical Committee ISO/TC204 based on eight principles of the Organisation for Economic Co-operation and Development (OECD) and was published as "ISO 24100:2010 Intelligent transport systems – Basic principles for personal data protection in probe vehicle information services" [25]. The following probe vehicle information services are defined as personal data: contract registration information with probe data suppliers, communication IDs, passwords for certification, communication logs, and personal data included in probe data itself.

Europe

The collection, storage and processing of personal data is regulated by data protection laws. Europe is working on a clear policy on privacy and confidentiality that increased protections for citizens while increasing the availability of public data to external users. Data protection legislation attempts to create a balance between protecting individual privacy and allowing data use for purposes (such as research) that are in the public interest.

As of May 2018, with the entry into application of the General Data Protection Regulation (GDPR), there is one set of data protection rules for all companies operating in the EU, wherever they are based [26]. The GDPR aims at providing greater transparency and consistency, reducing costs for businesses and increasing consumer confidence. The GDPR was designed to harmonise data privacy laws across Europe, to protect and empower all European citizens' data privacy and to reshape the way organisations across Europe approach data privacy. It also addresses the export of personal data outside the European Union.

USA

Private information relates to any data emitted, collected, or stored about individuals. A key concept in privacy analysis is Personal Identifiable Information (PII) [27]. PII is any information that can be used to distinguish or trace an individual's identity. PII is not specific to any category of information or technology, each case and associated risks must be individually examined for context and the combination of data elements that are provided or obtainable. The US model regarding privacy protection is typically that industry self-regulates unless and until problems or needs emerge. At that time, new legislation is considered.

Japan

In Japan, there are two national laws that relate to privacy and the big data issue, which are planned to be revised within the decade of 2020. The Act on the Protection of Personal Information (Act No. 57 of May 30, 2003) protects the rights and interests of individual persons [28]. It provides the basic idea relating to the protection of personal information as well as rules for handling personal information by certain private entities to be observed.

The Act on the Protection of Personal Information held by Administrative Organs (Act No. 58 of May 30, 2003) protects the rights and interests of individuals while achieving proper and smooth administrative management by providing for the basic matters concerning the handling of personal information in administrative organs [29].

China

The Chinese government released in January 2018 the final version of a new national standard on personal information protection, the Information security technology - personal information security specification [30]. It lays out detailed new regulations for user consent, as well as how personal data is collected, stored, and shared.

Australia

In Australia, the Australian Privacy Principles [31] apply and they cover:
- the open and transparent management of personal information including having a privacy policy
- an individual having the option of transacting anonymously or using a pseudonym where practicable
- the collection of solicited personal information and receipt of unsolicited personal information including giving notice about collection
- how personal information can be used and disclosed (including overseas)
- maintaining the quality of personal information
- keeping personal information secure
- right for individuals to access and correct their personal information.

6.6.2 How to deal with data privacy

Privacy-preserving mechanisms are needed for big data, so that provenance information and data access policies are not lost, and they are as follows:
- A “privacy by design” approach should be adapted in the development and application of big data analytics. This should include implementing technical and organisational measures to address matters including data security, data minimisation, and data segregation.
- Organisations should use appropriate techniques to anonymise the personal data in their dataset(s) before analytics. Anonymisation and obfuscation of some data values can be used to protect sensitive information. For example, geographic location may be generalised to a village or a town rather than the exact coordinates.
- For data exchange there is a need to create supply chain contracts or codes of practice. The availability of public data and public distribution channels needs to be regulated in a legal framework, which should define objectives and responsibilities of both, the public and the private sector. Data publishing and distribution policy also needs to be established.
- A privacy impact assessment framework into big data processing activities should be prepared to help identify privacy risks and assess the necessity of a given project. The privacy impact assessment should involve input from all relevant parties.

- Ethical principles should be developed with the aim to build relationships of trust with the public which will assist compliance with data protection requirements.
- To enable probe data suppliers to provide data without undue concern, the strict observance of personal data protection laws should be complemented by the preparation of guidelines to be followed by stakeholders and the standardisation of design guidelines would be necessary for that purpose.

6.7. Security

Security is very important for both data quality and for protection. Big data frequently moves across boundaries including boundaries of groups, communities of interest, states, as well as international borders. Provenance, a component of veracity, addresses the problem of understanding the data's original source and what has been done with the data. As with all data, the security of big data sets may be compromised if they are accessed by people without permission or used inappropriately by rogue individuals within an organisation. A range of tools and procedures can be used to reduce the risk of data being accessed without permission. However, it is impossible to guarantee that data will be completely secure.

Security measures taken by the road network operators must include:

- clear organisational structures and procedures
- quality assurance systems
- IT and communications network security, including data encryption

The Directive (EU) 2016/114811 of the European Parliament and of the Council concerning measures for a high common level of security of network and information systems across the Union (hereinafter: "NIS Directive") was adopted on 6 July 2016 [32]. According to the Directive Member States of the EU shall nominate operators of essential services, such as road operators. These operators shall therefore set up appropriate measures in order to ensure cyber-secured systems and services. Such measures are provided for instance by the ISO/IEC 27001, which is also the basis for recommendations of the ENISA (European Union Agency for Network and Information Security).

A variety of tools and procedures can support road network operators to make data processing more secure:

- Implementing effective data governance processes across an organisation (and external partners) can help to control access and use of data. This may involve making individuals accountable for data security, providing relevant training, minimising the number of people with access, and might result sometimes even in deletion of data, when appropriate.
- Encryption is a widly used tool for making access to data more secure. For example, a message can be translated into an unreadable code for transmission, and then later converted back into a readable form using a decryption key.
- It is not always necessary to exchange all data between all involved parties. Multiple parties can contribute data to a communal activity without having to exchange data with each other directly.
- Access to sensitive data may be provided via a "safe haven". This can be a secure location (cloud) or a set of administrative arrangements to assist safe and secure communication of confidential data.

- Prepare a security planning processes within the organisation. Development of a security plan provides effective means to meet cost-benefit and competitive resource challenges. Cybersecurity planning should incorporate, at the minimum.
- A security strategy needs to express management's commitment to cybersecurity and to provide the high-level direction and requirements for cybersecurity within the company.
- A security policy that address the range of management, personnel, operational and technical issues and guide the development, implementation and enforcement of the companys' security measures.
- Definition of roles and responsibilities that clarifies the decision-making party as well as the responsibility for cybersecurity.
- Vulnerability and risk assessments need to be set up to identify the company-specific security requirements and assist in prioritisation of risk management efforts.
- The development and maintenance of cybersecurity plans need to include risk mitigation and management as well as response and recovery plans.
- Active monitoring and evaluation needs to be done on a continuous basis.
- Awareness and training for all company employees needs to be ensured.

When planning for cybersecurity, this must be done in a systematic way by taking a balanced approach. It needs to be ensured that both, the security policy and controls, must be adaptable as failures are bound to happen.

7. CONCLUSIONS

Today, data is regarded as one of most important resources and big data will be the cornerstone input for better road network operations. Each road network operator uses data to improve planning, operation, and maintenance of the infrastructure they are responsible for. In general, agency owned sensor data or data generated by third parties (e.g. weather data) are used to support such activities of their daily works. Current trends show that more and more data are supporting the effectivity and efficiency of RNO services. These services might be used by the road network operator to improve internal processes, or these services might support road users.

When dealing with data, the general principles will apply to low- and medium-income countries as they do with high-income countries. Big data does not necessarily mean high investments into own transport infrastructure. Data can be generated within an organisation or it can be provided by third parties. It is however important that whoever generates transport related data, the ownership needs to be clarified. As a minimum, a road network operator should own the road infrastructure-related data and does not solely rely on third party sources. This especially critical as data ownership is key to allow intensive processing of small and big data coming from different sources, which will result in the most value.

In the case of big data, it was clear from the review that road network operators are either already inside the big data world or are close to become part of the big data world. Therefore, the hypothesis of this report that – big data is a groundbreaking tool for infrastructure operators, used to improve services for the achievement of policy goals and road users' expectations – can be fully supported. Data in general and big data especially has the potential to improve the design as well as the operation tasks of road network operators. Based on this fact, it is advisable to all infrastructure network operators to evaluate how and where big data can support their tasks. Ideally, big data should be part of an RNO strategy to be prepared for the future.

By setting up big data it is important that the contents of this report are considered. Awareness of the strengths as well as the weaknesses of big data will ensure that big data is used most effectively. It should also be understood that not all big data analysis will lead immediately to the expected results as oftentimes time is required to mature and refine applications. Therefore, if big data is not properly included as part of the organisational strategy, frustration due to lack of early results might lead to early cancelation of budgets needed for big data analysis. To set-up a big data strategy for road network operators, the process model described in chapter 5.3 could be used as a reference.

When thinking about a big data strategy, it is recommended to start with what is readily available and focus on the low hanging fruits. As stated, road network operators are already collecting lots of data for their daily work and merging this own data with data of third parties or other sources could be a first step towards the big data world. It is not necessary to rebuild existing and already operational systems. Nevertheless, initial adoption of collecting data and setting up systems in preparation of increasingly fast incoming and high-volume data (data lakes) should be considered. It is important to use these initial applications as the starting point for the integration of big data analysis. Metadata would help to make data accessible and it is an important aspect of a big data strategy. A standardised description of data as well as an identification of data ownership and conditions of usage is important. It is essential that metadata standards for transport are established and adopted by public and private sector data providers.

One of the most obvious benefit of big data is in the calibration of traffic models. Each road network operator has based their traffic management system including the internal decision support system and the external information strategy on traffic models. Big data will enable them to calibrate these models and to learn, if they work properly. A faster reaction on events or even prevention of incidents along the road network will be enabled by big data analytics.

Each road network operator should appreciate that big data analysis and knowledge about big data procedures will become a prerequisite for the next generation of traffic management and road network operations. Hence, big data skills need to be developed within the road network operator's organisation, either through internal or external sources. As road network operators are traditionally focused on civil engineering, a change in thinking is required which is sometimes only possible over a certain period. IT and related applications will play an increasingly important role in all core businesses of road network operation. If a road network operator is not in the position to run big data analysis by his or her own, then they may enable big data processing by making operator's data available to third parties. Private service operators might be willing to use RNO's data to process them with data coming from other sources and to provide services that might be of benefit to the road user. It is possible that other partiers other than the road network operator could provide the services that could make roads safer and more efficient.

As illustrated throughout in the report, big data analysis is in most cases done by bringing data sets of different nature and ownership together to develop new knowledge. To do so, cooperation is key. Visualisation of traffic patterns and better understanding of traffic can only be achieved when different stakeholders work together on the generation of new knowledge. The utmost knowledge cannot be developed within a single organisation, but only through the cooperation of various organisations, which includes public and private entities. A public road network operator needs to be aware of the legal aspects of big data particularly being open, neutral and especially discrimination-free towards different private entities.

The legal aspects of big data emphasise privacy and ownership regulations. Ideally, all data generated and used by road network operators will follow a "privacy by design" principle. Hereby all personal information within a dataset will be deleted automatically (e.g. at a roadside station) before any processing is performed (i.e. even outside the typical big data processing).

8. KEY RECOMMENDATIONS

The following key recommendations have been identified by working group B.1.3 of the Technical Committee B.1 Road Network operations/Intelligent Transport Systems for dealing with data in general and with big data specifically.

- When dealing with data, the general principles similalrly apply to low and medium income countries as they do with high income countries.
- Big data does not necessarily require big investments in own infrastructure. Big data can be collected by 3rd parties as described in the PIARC Technical Committee B1 Low Cost ITS report.
- A core element in big data is data ownership. It is recommended that road network operators take ownership on their own data, even when collected by 3rd parties, and not to buy data back from private companies.
- Data in general and big data especially has the potential to improve the design and operation tasks of road network operators.
- It is advisable to all infrastructure network operators to evaluate how and where big data can support their tasks. Ideally, big data should be part of an organisation-wide RNO strategy to be well prepared for the future.
- When developing a big data strategy, it is recommended to start with what is readily available and focus on the low hanging fruits.
- Not all big data analysis would immideiately lead to expected results as oftentimes time is needed to mature and refine applications. Therefore, if big data analysis is not properly embedded into the organisation’s strategy, frustration might lead to premature cancelation of budgets needed for big data analysis.
- The most obvious benefit of big data is in the calibration of traffic models.
- Big data skills need to be brought into the road network operator’s organisation, either as internal or as external sources.
- Cooperation in big data is key. Traffic patterns and chracteruitics can only be visualised and properly understood if various stakeholders work together.
- Follow a “privacy by design” principle.

9. GLOSSARY

TERM	DEFINITION	EXPLANATION
AI	Artificial Intelligence	AI is the development of machine learning systems that are able to perform tasks that usually require human intelligence.
ANPR	Automated Number Plate Recognition	
BIM	Building Information Modelling	Data is added to maintenance and service functions enabling infrastructure owners to develop a clear picture about the state of their assets, how they need to be managed, and what resources might be needed to preserve their capability.
BRT	Bus Rapid Transit	
BSS	Base Station Subsystem	
CAN	Controller Area Network	
CCAM	Cooperative Connected Automated Mobility	CCAM unifies current trends towards connected vehicles (also often referred to as cooperative systems or C-ITS) and automated road transport (including automated and autonomous vehicles).
CCTV	Closed Circuit Television	
C-ITS	Cooperative Intelligent Transport Systems	C-ITS is about connecting vehicles with vehicles, vehicles with infrastructure, and vehicles with everything.
CNG	Compressed Natural Gas	
DCAT-AP	Data Catalogue vocabulary (DCAT) – Application profile (AP)	The DCAT Application profile for data portals in Europe (DCAT-AP) is a specification based on the Data Catalogue vocabulary (DCAT) for describing public sector datasets in Europe.
DOT	Department of Transport	
DSRC	Dedicated Short Range Communications	
EV	Electric Vehicle	
ETC	Electronic Toll Collection	
FCD	Floating Car Data	
GDPR	General Data Protection Regulation	
GHz	Giga-Hertz	

TERM	DEFINITION	EXPLANATION
GPS	Global Positioning System	
GSM	Global System for Mobile communications	
GNSS	Global Navigation Satellite System	
HD map	High Definition map	High accurate digital description of the infrastructure that forms one basis for enabling CCAM.
ISO	International Standards Organisation	
IT	Information Technology	
IoT	Internet of Things	IoT is the concept of basically connecting any device with an on and off switch to the Internet (and/or to each other).
ITS	Intelligent Transport Systems	
KPI	Key Performance Indicator	
LIDAR	Light Detection and Ranging	
LNG	Liquefied Natural Gas	
MaaS	Mobility as a Service	In MaaS, a service operator provides tailor-made mobility services including all kinds of transport modes and covering all transport related matters from planning, booking and payments
MDM	Mobility Data Marketplace	
NAP	National Access Point	
NIST	National Institute of Standards & Technology of the United States	
OECD	Organisation for Economic Co-operation and Development	
OBU	On Board Unit	
OD	Origin – Destination	
PII	Personal Identifiable Information	
RCM	Rodos Crisis Management	
RNO	Road Network Operations	
SIM	Subscriber Identity Module	
SVO	Speed, Volume and Occupancy	

TERM	DEFINITION	EXPLANATION
TfL	Transport for London	
TMC	Traffic Management Centre	
USA	United States of America	

10. REFERENCES

[1] GOVERNMENT OF THE NETHERLANDS "Declaration of Amsterdam on cooperation in the field of connected and automated driving; Navigating to connected and automated vehicles on European roads", Netherlands, https://www.rijksoverheid.nl/documenten/rapporten/2016/04/29/declaration-of-amsterdam-cooperation-in-the-field-of-connected-and-automated-driving, 2016

[2] SCANIA "Scania takes lead with full-scale autonomous truck platoon", Sweden, https://www.scania.com/group/en/scania-takes-lead-with-full-scale-autonomous-truck-platoon/, 2017

[3] CEDR (Conference of European Directors of Roads) – MAASiFiE PROJECT RESULTS "The MaaS Ecosystem", Brussels, http://www.cedr.eu/strategic-plan-tasks/research/cedr-call-2014/call-2014-mobility/maasifie-project-results/

[4] EUROPEAN COMMISSION, "DCAT Application profile for data portals in Europe (DCAT-AP)", 2015, https://joinup.ec.europa.eu/release/dcat-ap-v11

[5] COPERATION AT, DE AND NL "Single Point of Access (SPA) – Coordinated Metadata Catalogue",2015, http://www.mobilitaetsdaten.gv.at/sites/default/files/pictures/SAP_Coordinated_Metadata_Catalogue_V2.0.pdf

[6] RODOS TRANSPORT SYSTEMS DEVELOPMENT CENTRE, "Czech Republic Highway Overview", https://rodos.vsb.cz

[7] C-ROADS - THE PLATFORM OF HARMONISED C-ITS DEPLOYMENT IN EUROPE, www.c-roads.eu, 2018

[8] CAR 2 CAR COMMUNICATION CONSORTIUM, www.car-2-car.org, 2018

[9] CHANDRASEKAR, P. "Big data and transport modelling : Opportunities and challenges", International Journal of Applied Engineering Research, 10(17), pp. 38038–38044., 2015.

[10] SEABORN, C., ATTANUCCI, J. AND WILSON, N. H. M. „Using Smart Card Fare Payment Data To Analyze Multi- Modal Public Transport Journeys in London Citation Accessed Citable Link Detailed Terms", Transportation Research Record: Journal of the Transportation Research Board, 2121.-1, pp. 55–62. doi: http://dx.doi.org/10.3141/2121-06, 2009

[11] TAO, S. ET AL. „Exploring Bus Rapid Transit passenger travel behaviour using big data", Applied Geography. Elsevier Ltd, 53, pp. 90–104. doi: 10.1016/j.apgeog.2014.06.008., 2014

[12] LEE, S. AND HUR, Y. "Night Bus: Route Design Using Big Data", Seoul Policies That Work: Transportation, pp.71-88, 2017

[13] MAZLOUMI, E., CURRIE, G. AND ROSE, G., „Using GPS data to gain insight into public transport travel time variability", Journal of Transportation, 136(July), pp. 623–632. doi: 10.1061/(ASCE)TE.1943-5436.0000126., 2009

[14] WORLD ROAD ASSOCIATION „Ride Sharing/Matching", https://rno-its.piarc.org/en/user-services-passenger-transport-transport-demand-management/ride-sharing-matching

[15] ARRB GROUP LTD AND AUTHORS 2016 "Network performance analysis for Perth congestion response", ARRB Conference, 27th, Melbourne, Victoria, Australia, https://trid.trb.org/view/1446648, 2016

[16] ESPADA, I. "Post-implementation assessment of pinch-point projects", 2017

[17] STATE OF VICTORIA "VicRoads Annual Report", https://www.vicroads.vic.gov.au, 2017

[18] MARGREITER, M. "Bluetooth technology for motorway management", 2018

[19] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, U.S. Department of Commerce "NIST Big Data Interoperability Framework: Volume 1, Definitions ", https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf, 2015

[20] GARTNER INC. (G00311517) Planning Guide for Data and Analytics, 13 October 2016, Analyst(s): John Hagerty, https://www.gartner.com/binaries/content/assets/events/keywords/catalyst/catus8/2017_planning_guide_for_data_analytics.pdf, 2017

[21] UEDA, S.(ITS-TEA Japan) "International standards for Low cost ITS and Big Data", PIARC: The 3rd TCB.1 meeting, Prague, April 2017

[22] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, U.S. Department of Commerce , “NIST Big Data Interoperability Framework: Volume 4, Security and Privacy”, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-4.pdf, 2015

[23] CLARKE, M. “Big Data in Transport”, The Institution of Engineering and Technology (IET), June www.theiet.org/sectors/transport/topics/intelligent-mobility/files/sector-insight.cfm, 2016

[24] UNITED NATIONS “Universal Declaration of Human Rights (1948)”, http://www.un.org/en/universal-declaration-human-rights/

[25] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION “ISO24100:2010 Intelligent transport systems – Basic principles for personal data protection in probe vehicle information services”, https://www.iso.org/standard/42017.html, 2010

[26] EUROPEAN PARLIAMENT “General Data Protection Regulation (GDPR) https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en, 2018

[27] U.S. DEPARTMENT OF LABOR “Guidance on the Protection of Personal Identifiable Information”, https://www.dol.gov/general/ppii, Washington, DC

[28] THE GOVERNMENT OF JAPAN “Act on the Protection of Personal Information (Act No. 58 of 2003)”, http://www.cas.go.jp/jp/seisaku/hourei/data/APPI.pdf

[29] THE GOVERNMENT OF JAPAN “Act on the Protection of Personal Information Held by Administrative Organs(Act No.57 of 2003)”, http://www.cas.go.jp/jp/seisaku/hourei/data/APPIHAO.pdf

[30] CHINA NATIONAL STANDARDS “GB/T 35273-2017 Information Technology – Personal Information Security Specification”, http://www.gbstandards.org/index/GB_standard.asp?id=106806&word=Information security technolog

[31] AUSTRALIAN GOVERNMENT – OFFICE OF THE AUSTRALIAN INFORMATION COMMISSIONER “Australian Privacy Principles”, https://www.oaic.gov.au/privacy-law/privacy-act/australian-privacy-principles

[32] THE PUBLICATIONS OFFICE OF THE EUROPEAN UNION (EU Publications) “Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union“, https://publications.europa.eu/en/publication-detail/-/publication/d2912aca-4d75-11e6-89bd-01aa75ed71a1/language-en

APPENDICES

10.1. ANNEX A: OVERVIEW OF METADATA ELEMENTS FOR A METADATA SET [5]

	Name of Metadata element	Field name (proposal)	Type of value	Field length (proposal)	Technical description	Example
Metadata information	Metadata Date	metadata_date	DateTime	-	YYYY-MM-DD'T'hh:mm:ssTZD; NOT NULL	2015-10-23T09:00:00+01:00
	Metadata language	md_language	Predefined Text	-	Predefined EU24 Language set ISO 639-2 conform; multiple choice; NOT NULL	ger; eng;
	Contact point for metadata	cp_name	Free text	50	Text; utf8; NULL	Hans Maier
		cp_org_name	Free text	50	Text; utf8; NOT NULL	Data GmbH
		cp_address	Free text	50	Text; utf8; NULL	Data street 1, Vienna
		cp_email	Free text	50	Text; utf8; NOT NULL	hans@data.at
		cp_website	Free text	50	Text; utf8; NULL	http://data.at
		cp_tel	Free text	50	Text; utf8; NULL	-
Content Information	Name of dataset	d_name	Free text	250	Text; utf8; NOT NULL	Highway network Austria
	Description of dataset	d_description	Free text	1000	Text; utf8; NOT NULL	Contains static high priority network of Austria, Link information: Speed, lanes, direction
	Dataset type category	data_agr_type	Predefined Text	-	Predefined 15 data categories; Lookup Table; multiple choice; NOT NULL	
	Dataset detailed type	data_org_type	Predefined Text	-	Predefined 50 data types; Lookup Table; multiple choice; NULL	

	Dataset language	ds_language	Predefined Text	-	Predefined EU24 Language set; single choice; NOT NULL	ger;
Temporal	Start date of publication	p_start_date	DateTime	-	YYYY-MM-DD'T'hh:mm:ssTZD; NOT NULL	2015-10-23T09:00:00+01:00
	End date of publication	p_end_date	DateTime	-	YYYY-MM-DD'T'hh:mm:ssTZD; NULL	2015-10-23T09:00:00+01:00
Geographical coverage	Area covered by publication	val_area	Predefined Text	-	Predefined NUTS 0 – 3 Codes; UTF8; multiple choice; NOT NULL	AUT11; AUT12;AUT13;
	Network coverage	net_category	Predefined Text	-	Predefined; UTF8; multiple choice; NOT NULL	Motorway
	Network coverage description	net_description	Free text	1000	Text; utf8; NULL	structural separated bidirectional lanes, 2 to 4 lanes, minimum speed 80, use condition
Responsibilitiers/ Contact Information	Publisher	p_name	Free text	50	Text; utf8; NULL	Hans Maier
		p_org_name	Free text	50	Text; utf8; NOT NULL	Data GmbH
		p_address	Free text	50	Text; utf8; NULL	Data street 1, Vienna
		p_email	Free text	50	Text; utf8; NOT NULL	hans@data.at
		p_website	Free text	50	Text; utf8; NULL	http://data.at
		p_tel	Free text	50	Text; utf8; NULL	-
	Data owner	do_name	Free text	50	Text; utf8; NULL	Hans Maier
		do_org_name	Free text	50	Text; utf8; NOT NULL	Data GmbH
		do_address	Free text	50	Text; utf8; NULL	Data street 1, Vienna
		do_email	Free text	50	Text; utf8; NOT NULL	hans@data.at
		do_website	Free text	50	Text; utf8; NULL	http://data.at
		do_tel	Free text	50	Text; utf8; NULL	-

Condition for use	Contract or licence	con_lic	Predefined Text	-	Predefined; UTF8; single choice; NOT NULL	Licence
	Condition for use	con_description	Free text	1000	Text; utf8; NULL	http://data.at/terms.pdf
Access information	Structure of publication	tech_structure	Predefined Text	-	Predefined; single choice; utf8; NOT NULL	Datex II XML
	Publication structure description	p_structure_description	Free text	1000	Text; utf8; NULL	http://data.at/tech_description.pdf
	Access interface	interface	Predefined Text	-	Predefined; single choice; utf8; NOT NULL	SOAP;
	Communication method	com_method	Predefined Text	-	Predefined; Multiple choice; utf8; NOT NULL	push;
	Access URL	access_url	Free text	250	Text; utf8; NOT NULL	http://data.at/access.csv
Quality information	Update frequency	update_freq	Predefined Text	-	Predefined; Single choice; utf8; NOT NULL	yearly
	Quality Indicator	qm_indicator	Free text	1000	Text; UTF8;NOT NULL	According to the EIP+ quality measures
	National Body assessment date	assessment_date	DateTime	-	YYYY-MM-DD'T'hh:mm:ssTZD; NULL	2015-10-23T09:00:00+01:00

10.2. ANNEX B: CASE STUDY CZECH REPUBLIC ON BIG DATA PROCESSING

By Marek Ščerba and Martin Hájek (Technical University of Ostrava)

Description

A consortium headquartered at the Technical University of Ostrava has developed an innovative “dynamic mobility model” that use a range of data to predict traffic flows and to enable traffic management decisions.

The consortium known as RODOS, uses as its raw material traffic movement data from company car fleet management databases (140 thousand equipped vehicles), all traffic detectors on the main road network and data on truck movements (ETC). Overlaid on this is weather information, creating a fusion of data sources. The data processing used the supercomputing centre in Ostrava. The working name of the system is called viaRODOS.

Output data describing current traffic flow dynamics contains the following parameters:

- current average traffic flow and speed in each segment of the road network in CR (45 000 segments);
- current travel time necessary to pass defined segments;
- current delay in defined segments as compared to travel time necessary to pass the segment with no obstructions, typical average speed for passing the segment with no obstructions;
- speed profile (average speed/delay) in a specific day and hour.

Objectives

The viaRODOS system was primarily designed and implemented as a key and nationwide monitoring system based on robust datasets. Their integration and data fusion were intended to be the basic tool for modern transport management. The goal was to demonstrate through the research the viability of the big data system and to verify the possibilities of efficient provision of traffic information with existing and new data sources as well as to verify the possibilities of integrating real-time data into traffic management systems.

Other objectives were to prove that the fusion of the data sources eliminates the deficiencies of the use of separate data source and, on the contrary, to ensure the effective ways of extracting the available data sources with mutual interconnection into one output with high accuracy and high service availability.

Technical challenges

The projects main technical challenge was the interconnection of input data, the robustness of the database, and the mutual complementarity of GPS sampling with the already installed profile detectors (ASIM), the data from the national electronic toll collection system and the data from the weather forecasts.

- To fuse elementary data at a low level - at the level of space-time transactions while meeting the anonymity requirements of each system - enables efficient combination of all data sources without loss of information by premature aggregation at aggregation intervals;
- to fuse GPS data with current data sources (ETC, Detectors) at a deep level - at the level of space-time transactions, while meeting the anonymity of each system;

- to adjust the fluency with the user interface of the system, respectively with database and system performance optimisation for working in real-time;
- collecting and setting up user's requirements for a native mobile application realisation.

Non-technical challenges

One of the non-technical challenges was organising the connection of various fleet monitoring service providers (electronic journey logs, anti-theft protection, etc.). Individual numbers of vehicles had to be linked to a single database. In this context, it was also necessary to ensure that commercial entities were willing to provide this data, despite limited financial resources from the research project. Key principles applied to address the challenges were as follows:

- cooperation with the Road and Motorway Directorate as part of the exchange of existing data sources and pilot testing of the system within the National Traffic Information Centre;
- discussion with stakeholders and creation of a project plan at the national level;
- creating awareness of the existing system with experts, but also with drivers;
- create a positive media image of the system;
- collaboration in creating conditions for data procurement at national level.

10.3. Annex C: Case Study on Automated Traffic Signal Performance Measures in the State of Utah

By Mark Taylor (Utah Department of Transportation)

Description

Automated Traffic Signal Performance Measures (ATSPMs) are a series of visual aids that display quantitative data from high-resolution data logs from traffic signal controllers. The data logs record nearly every action the signal controller makes (detectors on and off, phase state changes, controller events) and time-stamps the enumerations. ATSPMs are a valuable asset management tool, aiding technicians and managers in the control of both traffic signal hardware and traffic signal timing and coordination. They allow analysis of data collected 24 hours a day, 7 days a week, improving the accuracy, flexibility, and performance of signal equipment and the system as a whole. ATSPMs show real-time and historical functionality at signalised intersections. Engineers can use ATSPMs to directly measure what they previously could only estimate and model. ATSPMs allow agencies to effectively optimise and manage traffic signals without extensive field data collection, and to visualise data in an easy-to-understand format that is web-based and easily accessible by a diverse group of users. Using ATSPMs, agencies will be able to improve mobility, increase safety, and use resources more effectively.

Signal controllers with high-resolution data logging capability log controller events with a 1/10th resolution timestamp and store the events in temporary data files on the controller. A file transfer protocol (FTP) connection is made periodically (about every 15 minutes) from a central server to the traffic signal controller to retrieve the data files. Once collected, the data files are translated into a database, where they can be interpreted manually or with an automated graph, analysed visually or with an optimisation algorithm, and archived for comparative analysis. An efficient means of presenting the information is to use a web interface that allows all users – such as agency, consultants, academia, executive leaders, and public officials – the chance to view the information without requiring special software or needing to manage firewalls and network connections.

The Utah Department of Transportation (UDOT) started development for ATSPMs April 2012 with some assistance with Purdue University and the Indiana Department of Transportation. UDOT and others continues to develop and enhance ATSPMs. The ATSPM source code is shared with others through the open-source portals at GitHub and FHWA (OSADP).

Each intersection logs approximately 11 MB to 15 MB of data per day. With over 1800 intersections bringing back data and indexes created to assist in finding the records, UDOT servers log about 1 TB of data each month. From January 2014 to January 2018, UDOT has approximately 36 TB on its server currently in use.

Objectives

- Analysis of automated data collected 24 hours a day, 7 days a week.
- Creation of an independent web interface tool that allows all users such as agency, consultants, academia, executive leaders and the public a chance to view signal performance measures without requiring special software or needing to manage firewalls and network connection.
- Improve traffic signal operations by targeting and prioritising problem areas as well as optimising signal timing parameters.

- Improve traffic signal maintenance by knowing in real-time equipment malfunctions such as vehicle and pedestrian detection not working properly.
- Assist other agencies in the transportation industry in the development, integration and use of ATSPMs.

Technical challenges

- The ATSPM source code was re-written so all metrics and the database could be under a single architecture. The software re-write allowed it so the source code could be distributed on the FHWA's Open Source Application Portal (OSADP) at: https://www.itsforge.net/index.php/community/explore-applications#/30/133 and GitHub at: https://github.com/udotdevelopment/ATSPM so others outside of UDOT can contribute and partner in the further development of ATSPMs. It took UDOT developers about 5 months to re-write the source code in a more stable single architecture.
- An FTP program was written to gather all records from all intersections and a record translator (data logs on signal controller are in binary format) and a partitioning scheme in the database was developed so all records from all signal controllers can be retrieved in a timely manner every 15 minutes. It was necessary to create these two items carefully so translation of the records into the database did not fall behind the 15-minute schedule and bog down the system.
- A distributed architecture was created so the processing components are on a different server than the web components.

Non-technical challenges

- ATSPMs have required several years and months of development by internal software developers that work for the Department of Technology Services at the State of Utah. They are non-UDOT employees. A good business case was presented and priorities were made in allowing the State of Utah software developers to spend time on this UDOT project.
- UDOT is overseen by a Transportation Commission who is appointed by the governor and serve as part of an independent advisory committee. The group prioritises projects and decides how funds are spent. One of the commissioners disagreed with UDOT's decision of donating the ATSPM source code to open source and not selling it to make a profit for all the software development that went into it.
- Permission was needed from our State Attorney General's office and State of Utah Department of Technology Services for the ATSPM source code to be donated to open source and to other agencies.

10.4. ANNEX D: CASE STUDY NETHERLANDS ON REDUCING RUSH HOUR CONGESTION

By Emmanuelle Frénéat and Elena Umanets (EGIS)

Description

The Avoiding Rush Hour (ARH) initiative aims at encouraging commuters not to drive during peaks hours by rewarding participants (via cash, credit on public transport cards, loyalty points to be spent in a web shop, gamification elements, compensating parking costs) when they don't drive during rush hours within a defined network or corridor.

The Avoiding Rush Hour program has been tested since 2011 in Rotterdam (Netherlands) with 12,000 participants. The program awards a personal budget to every participant calibrated on his/her average number of trips during rush hours before entering the program.

The system, developed in Rotterdam, brings together 12,000 participants. Tested for over five years, the initiative seems to be proving its effectiveness, reducing traffic by between 5 and 10% during peak hours. Thanks to this system, 4100 journeys are now avoided every day in Rotterdam, with an average of 40% participation per day for €3 credit in cash or €3.50 on a transport card.

Objectives

Traffic congestion is currently a significant barrier to the economic development of large and medium-sized cities, with an average of 40% of time lost on busy roads over a one-hour journey. This lost time represents a significant cost for the community (fuel, working time wasted) and contributes to making regions less attractive. The maintenance costs of these roads, congested in the mornings and evenings, are high. This rush hour congestion has a negative impact a) on quality of life as it increases daily stress, b) on air quality through emissions of harmful gases and microparticles and c) on environmental quality, because of greenhouse gas emissions that contribute to global warming.

The analysis of the factors leading to this suburban congestion highlights long-established organisational and societal models and deep-rooted individual behavioural patterns: synchronous morning commutes, concurrence of driving times for individuals and other transportation, the phenomenon of single occupancy car use, longer distance commuting, low use of public transport, door-to-door car journeys, and so on.

Two alternatives are traditionally implemented: 1) an increase in road capacity and public transport and 2) the regulation of traffic using tolls or a tax.

In 2008, a "Spits Mijden" or rush hour avoidance scheme emerged in the Netherlands It aims to reward users of road infrastructure for its "non-use" during peak hours and thus work towards reducing the intensity of traffic. This innovative approach has led to monetary reward schemes for drivers not using their cars during peak hours.

Technical challenges

Participants are rewarded for not driving during peak hours within a defined network or corridor. To challenge the participant a personal budget, based on their personal reference level, is awarded to every participant. This personal reference level reflects the average number of trips during rush hours before entering the ARH project.

To define this reference level, ANPR cameras have been installed along the relevant highway prior to the start of the project. During a period of 4 to 6 weeks all license plates have been registered. Based on these registrations the reference level per license plate has been defined (for example 4 registered trips during a week). All measurements have been anonymous. The license plates with the minimum required reference level or higher will later be linked to people through the national register of license plates, creating a database of potential participants. All potential participants will receive a personalised invitation to join the project. Based on the participant´s reference a personal budget is offered. For example, when an avoidance is valued at € 3.50 a participant with reference level 4 will have a weekly budget of € 14.

If the participant accepts the conditions he can join the program. By accepting the conditions the participant also agrees to his driving behaviour being monitored (on the defined corridor or network only). Every trip during rush hours will result in reducing his available budget with € 3.50. If in the above example the participant does take a trip during rush hours his total reward for that week will be € 10, 50, which will be transferred into the participant's bank account at the end of that week. The following week the participant will again start with a total budget of € 14,-.

On average, with a minimum reward level of 2.5 euros per avoided or deferred journey, the rate of active participation observed has been 33%. This allows an 8 to 10% reduction in traffic, which is enough to unclog a motorway at rush hour. From a technological point of view, ANPR cameras are used to identify motorists and ensure control of fraud. The technology is reliable and effective, but its use requires some technical precautions (no data retention, anonymity, etc.) and the submission of files to the relevant authorities.

Non-technical challenges

The main non-technical challenges refer to privacy issues.

For defining the reference levels and monitoring driving behaviour, different solutions can be used: ANPR, Apps or an On-Board Unit. There are also different ways to reward behavioural change such as cash, credit on public transport cards, loyalty points to be spent in a web shop, gamification elements, compensating parking costs, etc.

All these types of projects require special attention to privacy. All operators have to be registered at the privacy council and act accordingly. This means removing data as soon as possible (in all phases of the project) and not sharing any personalised data with external parties. For sharing their data with the operator, the participants will be compensated when rewarded for avoiding rush hours. Based on these principles these kinds of projects are very well accepted.

10.5. Annex E: Case Study Australia (Sydney) on post-implementation analysis

By Ian Espada (ARRB)

Description

The Pinch Point Program (PPP) of the Roads and Maritime Services (RMS) targets traffic congestion points, intersections or short lengths of road at which a traffic bottleneck exists (RMS 2015). The program aims to alleviate locations where traffic can build up resulting in travel delays at both the location and surrounding network.

The original PPP ran for five years from 1 July 2007 to 30 June 2012. A commitment was made to extend the program. In a typical financial year, eleven projects are constructed with 25 proposals developed.

Pinch point projects can vary depending on the problem they are addressing. Example projects include improving intersections, widening small sections of roadway, lengthening or widening busy turn bays, replacing heavily-used roundabouts with traffic signal, implementing new tidal flow traffic schemes, implementing clearways and parking restrictions during peak traffic periods and incident management such as installing closed circuit television (CCTV) and variable message signs (VMS).

The benefits of the pinch point projects are wide but primarily include improved travel time, travel time reliability, productivity and operational efficiency. However, many of these benefits are not quantified as RMS does not have access to an assessment process, methodology, and tool which is easy to use and interpret the results; not require significant lead time for the purpose of data collection; and can be used to assess any pinch point project (not just signalised intersections with traffic signal loop detector data, i.e. SCATS data).

The methodology developed for the PPP involved the use of traffic count data and travel speed data. A tool was developed referred to as Post-implementation Traffic Network Assessment (PITNA). PITNA utilises SCATS traffic count data and travel time data from probes to calculate performance indicators. Probe data can be sourced from RMS or other data suppliers. The development and pilot of PITNA utilised the RMS probe data sets. Probe data sets are travel speed data from vehicles equipped with GPS devices or other technologies that can detect location, such as smartphones. These devices communicate their location and speed which is then utilised to develop a travel speed database of the network. The travel speed database from probes is of high temporal and geographic resolution. The percentage of vehicles on a highway segment with probe devices is relatively low, but over a long period the number of samples that can be generated would be adequate for analysis in most cases.

PITNA was initially applied to two PPP projects, which included the Pacific Highway Off-ramp Widening project and the Victoria Road Weekend Clearway project. The application of PITNA clearly demonstrated positive impacts of the two projects. The assessment was later expanded to cover nine more projects of various types. A range of impacts were identified, but overall it was clear from the evidence and analysis that the pinch-point program was delivering significant benefits in terms of relieving congestion delay. The PITNA results contributed to the assessment of the program.

Objectives

The purpose of this project was to develop a process, methodology and tool to assess the benefits of pinch point projects. Key specifications of the project outputs are as follows:

- The methodology is simple, easy to understand, robust, inexpensive and able to be used for all relevant existing RMS datasets.
- The tool will be developed using a MS Excel.
- Data collected for the assessment can be derived from traffic and road transport systems, such as SCATS and probe vehicle fleet data, but will remain independent of a system to be adopted.

This project was limited to the assessment of operational improvements and excludes impacts on safety or travel behaviour (e.g. mode share).

Technical challenges

Key technical challenges included the following:

- Certain roads, particularly side roads, sometimes do not have enough probe speed data. This prevented the full analysis of impacts for certain projects.
- The quality and reliability of traffic loop detector data was a key constraint. Changes in detector labels and sometimes breakdown in equipment impact the quality of analysis.
- Projects involving roundabouts are not equipped with loop detectors, hence traffic count data was not available which constrained the quality of analysis.

10.6. Annex F: Case Study London Orbital (UK) on the use of near real time dashboards to improve operations performance on the M25 London orbital motorway

By Emmanuelle Frénéat and Erwann Huerre (EGIS)

Description

Like most organisations, infrastructure operators are generating large amounts of data and are looking for new means of turning this data into useful information to improve their efficiency and the quality of the service they deliver to their clients. Analysis of big data is a means of achieving both contractual compliance and improved performance by harnessing information captured by operational systems.

This principle has been successfully implemented on the M25 DBFO (Design Build Finance & Operate) project in the UK. This project, which runs the London orbital road, one of the most heavily trafficked motorways in Europe, is using a combination of information from various data sources and systems to produce real time or near real time alerts and information dashboards.

All data collected through operational systems across the network is centrally accessible via a 'data warehouse' and used to produce near real-time performance dashboards, which provide automated early warnings of any significant changes in operational performance data and allow managers to put in place the required improvement actions (typically in areas such as: incident response, defect management and winter maintenance).

These dashboards are accessible through the company intranet and allow users to select specific research criteria or to drill down various elements of the dashboard to access the supporting data, thus allowing quick and easy access to the required information. These interactive dashboards can also be issued automatically via email to identified recipients at set intervals.

Objectives

In a live environment such as a motorway, where life critical decisions have to be taken very quickly, real time or near time computing is a powerful tool to ensure that an event, such as an accident, can be managed within a specified time constraint. The constraint can be defined by the business user in accordance with the IT system and architecture.

On the M25 network digital cameras are streaming live video 24/7. These pictures are monitored in the Network Operation Centre (NOC) and allow an operator to react in real-time to an incident. The operator notifies the closest Incident Support Unit (ISU) of the incident and location. Details of the incident are recorded in an integrated incident and road management system.

Data from the management system flows into a data processing solution referred to as the Data Warehouse on the back of processes designed to run at specified intervals. The data is prepared for reporting in accordance with a defined set of rules to test the quality and consistency of the information received.

A performance dashboard delivers the incident details to a target audience via an auto generated email or text message. The information is also published on the intranet for a wider audience to analyse. Early notification of the incident is generated in near real time to help make informed decisions on live operational activities.

The information flowing through the Data Warehouse can also be used 'off-line' to produce performance reports or dashboards. Performance dashboards are interactive reports that allow users to identify high-level trends. The details in the big picture can then be easily accessed by using a drill through method, thus allowing managers to quickly identify the root causes of any negative performance pattern to address them. Information can be rendered in common formats, Word, Excel and PDF, etc. and delivered typically via an auto-generated email or published on Microsoft SharePoint, the content management system.

Technical challenges

The Data Warehouse manages processes that collect data from across the M25 network, which effectively includes any electronic data generated by the operation, whatever its format and location on the network. A process or job is designed to run automatically at specified intervals and is cumulative; meaning that a new set of information is added each time the job is run. This type of data can be used for reports that show metrics, trends, aggregations, and other relationships among the data. Data is cleansed, validated and formatted before it is used to feed the business and contractual reporting requirements.

Any failure of the information system on the M25 could have an immediate adverse impact on performance. In the case of incident management, any delay in the response could result in secondary accidents or important congestion.

To add resilience to a process, the Data Warehouse is hosted in a clustered SQL Server environment, a group of servers that support server applications with a minimum of downtime. Inactive servers within a group or cluster, provide continued service when there is failure on an active node. The SQL cluster is part of a data centre that is delivering IT Services to the project.

Non-technical challenges

The main non-technical challenges lie in the operatives' ability to express their needs so that the reporting tool serve the right purpose.

10.7. Annex G: Case Study Japan on Electronic Toll Collection

By Yuji Ikeda (MLIT)

Description

Japan has been working on development and introduction of the "ETC2.0" system as a vehicle-infrastructure cooperative ITS since 2014. The ETC2.0 system adopts DSRC in the 5.8GHz band that allows high-speed and high-capacity communications between vehicle and infrastructure. In this system, in addition to the basic applications such as ETC (Electronic Toll Collection), safe driving support information provision, and dynamic route guidance, it is possible to collect probe data from the vehicles equipped with the special on-board units (OBUs) with GPS.

Probe data collected by the ETC2.0 system is called as "ETC2.0 Probe Data" and being recognised as important "big data" in transport sector in Japan.

By analysing the ETC2.0 Probe Data, it is possible to identify bottlenecks and locations where sudden braking/steering occur frequently, and thus to clarify if there are some problems in road traffic. In Japan, a number of efforts have been made such as site-specific congestion measures and safety measures based on such analyses.

Objectives

In recent years, basic policies in Japan road administration emphasises the importance of utilising the existing road network infrastructure more efficiently.

In order to manage road network more efficiently, it is crucial to

- observe actual traffic conditions,
- identify problems through the analysis of the observed data, and,
- plan, implement, and evaluate the measures to solve the identified problems.

It is expected that by using big data such as ETC2.0 Probe Data when identifying problems and conduct ex-post evaluation, we can facilitate the project prioritisation and more efficient operation of road assets.

Technical challenges

Outline of the ETC2.0 Probe Data collection

The ETC2.0 Probe Data is collected from the vehicles' OBUs to the roadside units (RSUs), which are managed by road administrators. The data is sent from the RSUs to the "Probe Server", then aggregated and stored in the server, finally provided to road administrators. Road administrators use this data for providing drivers with road traffic information (wide-area information on congestion, travel time information, etc.) and safe driving support information, as well as for R&D activities and road management. However, sending probe data from OBU to RSU is arbitrarily, which means drivers can choose either they send the data or not.

Outline of the ETC2.0 Probe Data

The ETC2.0 Probe Data consists of a) basic information, b) travel records, and c) behaviour records. Here, considering the privacy, the unique IDs of OBUs are converted to irreversible ones on the probe server before data processing. Thus, it is impossible to identify individual vehicles or drivers from the ETC2.0 Probe Data.

- Basic Information

Basic information consists of the information about the OBUs (wireless unit [manufacturer, model number, etc.] and car navigation system [manufacturer, model number, etc.]) as well as the information about the vehicle.

- Travel Record

Travel record includes time stamp, coordinate (latitude and longitude), and road category (expressway, urban expressway, arterial road, or other). This data is recorded every 200-m driven and when vehicle turns 45 degree or more. Here, for privacy purposes, the OBUs delete the data near origins and destinations (for example, approximately 500m before and after turning on/off the engines). The storage capacity of an OBU is approximately 80 km of the data in maximum.

- Behavior Record

Behavior record includes time stamp, coordinate (latitude and longitude), direction, road category, longitudinal acceleration, lateral acceleration, yaw angle velocity, etc. This data is for the sudden behavior of vehicles by which either of longitudinal acceleration, lateral acceleration, or yaw angular velocity exceeds the predetermined thresholds. Among the consecutive sudden behavior, the data is recorded only at the position with the highest peak values. The sudden behavior that exceeds the thresholds indicates the possibility that drivers tried to avoid some dangerous situation. The storage capacity of an OBU is 31 events for this behavior record.

In addition, the probe server aggregates the travel record of individual vehicles and provide the data in a straightforward format such as travel time, travel speed in certain time units and section units (applicable to the digital road map), in order for road administrators to easily analyse and understand traffic situations.

Non-technical challenges

In order to ensure the data accuracy, the size of ETC2.0 Probe Data has to be large enough.

For that, promoting the use of the OBU is important. The number of the OBUs equipped vehicles has kept increasing since the launch of the ETC2.0 system in 2014 and was about 2.9 million as of June 2018. The percentage of the vehicles using the ETC2.0-OBUs was 17.3% in June 2018.

As the ETC2.0 Probe Data is collected by the RSUs, in areas with RSUs installed scarcely, data collection may be delayed or data accumulation may exceed the storage capacity of the OBUs. Currently, approximately 1,700 RSUs are installed on expressways (Figure 4), and approximately 1,900 on national highways under jurisdiction of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT).

10.8. Annex H: Case Studies South Korea on Public Transport Network Planning

By Keechoo Choi and Yongju Yi (Ajou University)

Description

The bus system is one of the most popular transit modes for both the city and the country. Due to its flexibility of routes operation and excellent accessibility, bus is a major transit mode in most cities. In Seoul, the capital of South Korea with a population of 10 million, the modal share of bus transit is 26.1% (as of 2016), in spite of the existing dense subway network. Meanwhile, in Suwon, a suburban city outside of Seoul that has a population of 1.2 million, the modal share of bus transit is 34.8% in an area with an insufficient number of subway lines.

In recent years, however, the modal share of bus transit in Korea has been stagnant or slightly decreasing due to the increase in passenger car ownership and the expansion of road and subway networks. To improve the competitiveness of bus transit, express bus service has been introduced and is an emerging form of urban transit service as an alternative to improve the competitiveness of urban public transit. For bus system to be a successful system, the careful selection of bus stop, so called here EBSS (Express Bus Stop Selection) is important. In this study, a model and a method that minimise a user's total travel time or total social costs with using bus smart card data has been proposed. More specifically, this study 1) develops the EBSS model considering the practical constraints using route and travel data obtained from smart card data, and 2) provides a case study on Suwon city to verify and evaluate the EBSS model. The method set forth in this study can be utilised as a tool to easily compare the operating efficiency between express bus and normal bus systems.

Objectives

- Optimising the express bus service with considering both users and operators
- Improve competitiveness of public transit on a limited budget
- Improve bus network efficiency without large-scale renovation while most bus routes are adjusted based on smart card data and civil complaint
- Increasing mobility of transit without decreasing current accessibility

Technical challenges

Use of Genetic Algorithm for Practical Application

- Genetic algorithm is one of the typical global optimum solution finding methods. Therefore, it is widely used in TRNDP (Transit Route Network Design Problem); Applying Heuristic methods because of model complexity and huge scale of network
- Solution size is same as the number of stops (stations), so dimension of solution is quite high. (each gene has binary integer whether express-bus stop or not at certain station)
- Expandability for applying real network: Huge size, High complexity
- Expandability for Further research: Applying passenger arriving probability, combined model for route and stops selection model, etc.

Smart Card Data Application (in Korea)

- After the implementation of distance-based integrated transit fare for the Seoul metropolitan area in 2007, smart card usage increased greatly – up to 98 percent – because of the implemented discount on transfer trips. In Korea, all data includes temporal and spatial references for all transit modes which are used in one trip; this data can be gathered due to the tagging smart card both when boarding and alighting.
- Data including User class (ordinary, students, senior, disabled, …), Mode codes (bus, rail, …), Route ID and Vehicle ID, Total distance(m), Total In-veh time and Total Time (min), Boarding and alighting time, Boarding and alighting stop ID, Total cost
- Actual route and trip data from smart card data are used to extract representative values of each factor that influences the effectiveness of the express bus implementation. Stop-based ODs by each route are created by each passenger's boarding/alighting time and stop information.

Travel Time Reduction by Skipping Local-only Stops

- Acceleration + Deceleration time saving (11.6 sec per stop)
- Boarding and alighting time saving (Boarding: 2.3sec/passenger, Alighting: 2.0sec/passenger (calculated from smartcard data))
- Intersection signal delay saving: Reducing expected delay time (go through green light without stopping) caused by arriving intersection faster as amount of "Acceleration + Deceleration time saving" and "Boarding and alighting time saving"

Non-technical challenges

While most of the existing research related to express transit services deals with restructuring the whole network or improving a single route, most of the cities have quite well-organised transit route networks through spontaneous development or practical policies to satisfy users' needs and operators' efficiency. Therefore, instead of restructuring the whole transit network, this study provides the methodology to select an optimal route set from the currently operating transit network that is maximising system efficiency without large-scale restructuring of the whole network.

World Road Association (PIARC)

La Grande Arche, Paroi Sud, 5e étage, F-92055 La Défense cedex

ISBN 978-2-84060-559-1