LEVERAGING BIG DATA FOR MANAGING TRANSPORT OPERATIONS

Deliverable 2.1 **Report on Economic and Political Issues**

Tharsis Teoh[1], Ivo Hindriks[1], Kim Hee[2], Bastiaan van Berne[1], Paul op 't Ende[1]

Panteia B.V.[1], Goethe-University Frankfurt[2]

August 2018

Work Package 2

Project Coordinator

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

Horizon 2020 Research and Innovation Programme

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

Distribution level	Public (P)
Due date	31/07/2018 (M10)
Sent to coordinator	28/08/2018
No. of document	D2.1
Title	*Report on Economic and Political Issues*
Status & Version	*Final*
Work Package	*2: Institutional and governmental issues and barriers*
Related Deliverables	*D2.2, D2.3, D2.4*
Leading Partner	*Panteia*
Leading Authors	*Tharsis Teoh, Panteia (Chapter 2,3,4)* *Ivo Hindriks, Panteia (Chapter 1,4,5,6,7,8)*
Contributors	*Kim Hee, GUF (Chapter 4)* *Bastiaan van Berne, Panteia (Chapter 4)* *Paul op 't Ende, Panteia (Chapter 7)* *Julien Debussche, Bird&Bird* *Jasmien César, Bird&Bird*
Reviewers	*Rémy Russotto, CORTE* *Marusa Benkic, CORTE*
Keywords	*Big Data, Transportation, Economy, Politics*

Disclaimer:

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

The content of this report reflects only the authors' view. The European Commission and Innovation and Networks Executive Agency (INEA) are not responsible for any use that may be made of the information it contains.

Executive summary

The proliferation of data availability is reshaping the economic and political realm. On the one hand, big data enables private and public parties to provide better quality services and products. On the other, the usage of data has led to policy response for limiting (e.g. the GDPR) or enabling (e.g. EU's policy towards a 'digital single market') the application of big data.

This report aims at revealing the wider economic and political issues involved with utilizing big data by elaborating on the interaction between transport actors (demand-, supply-, external- and governance actors) and their role in the data economy (as data users, suppliers or facilitators). Subsequently, the interaction between these actors is described on various levels.

On the **firm level**, private parties use big data use big data for improved situational awareness of the transport system, improving the capacity of transport networks, improving transportation services and facilitating the shift to sustainable transport.

On the **industry level**, the data economy is expected to grow rapidly in the coming years. While the EU data economy remains to be in a deficit compared to the US regarding structural factors (fewer data SME), cultural/educational factors (ability to create and keep data-related skills), and the presence of IT giants, there is a healthy presence of digital start-ups and innovation capacity.

On the **national level**, governments utilize big data in improving organisational performance and in service provision and policy making. Subsequently, big data is applied in transport related government tasks, including transport planning, traffic monitoring and public transport provision.

On **international level**, governments want to control data flows to limit the negative consequences of data, e.g. preventing the misuse of personal or classified data. Another reason is to easier carry out their task as supervisor, e.g. demanding local storage of tax or gambling data to simplify control routines.

Having discussed this, the most critical challenges for the future identified in this report are:

- Lack of data professionals, in particular within governments;
- Government compartmentalization, limiting optimal data usage by governments;
- An insufficient framework that satisfies both the user demand for privacy and the usage of personal information for business innovation. This is particularly relevant for public-private data sharing schemes;
- Too little awareness on the capacity of big data, as 'bad' big data analysis happens quickly, i.e. too little knowledge on what transport-related questions it can and cannot address, and how big data should address these questions.

This document is valuable in the objective definition stage of big data applications in transport. It facilitates the discussion on *if* big data should be used by providing insight into how that may be. Economic actors are provided with an overview of existing applications of big data in the transport sector. Political actors are given insight into how governments are currently using big data.

Table of contents

Executive summary II

List of Figures VII

List of Tables VIII

Glossary IX

1 Introduction 1

1.1 Abstract.. 1

1.2 Purpose of the document.. 2

1.3 Target audience 2

2 Framework for Analysis 4

2.1 Stakeholders in the transport sector.. 4

2.1.1 Transport demand 4

2.1.2 Transport supply 5

2.1.3 Transport externalities.. 6

2.1.4 Transport governance.. 6

2.2 Role of big data in the transport sector 7

2.3 Dimensions of analysis 9

2.3.1 Economic dimensions.. 9

2.3.2 Political dimensions 10

2.3.3 Highlighted issues: privacy and market power.. 11

2.4 Summary.. 11

3 Firm Level: Influence of Big Data on Business Models of Transport Companies.. 12

3.1 Direct improvements to the transport services 12

3.1.1 Shared situational awareness of the transport network 12

3.1.2 Improving transport network capacity for all modes 12

3.1.2.1 Long term planning of the transport network.. 12

3.1.2.2 Real-time traffic management 13

3.1.2.3 Mitigating risks that degrade network capacity 14

3.1.3 Improvement of transport services.. 15

3.1.3.1 Marketing activities in the transport sector 15
3.1.3.2 Vehicle routing 16
3.1.3.3 Service provision and schedules 16
3.1.4 Shift towards sustainable mobility and transport 17
3.2 Role as a data supplier 19
3.3 Costs of innovation 20
3.3.1 Bad or dark data 20
3.3.2 Data acquisition cost 21
3.3.3 Systems acquisition and operation 22
3.3.4 Re-organization and adaptation of processes 22
3.4 Summary 23
4 Industry Level: Growth of an Industry using Big Data Technology 25
4.1 Actors: Growth of data suppliers, data users and the data economy 25
4.1.1 Data companies: suppliers and users 25
4.1.2 Data workers 26
4.1.3 The Data Market and the Data Economy 26
4.1.4 Other data markets: USA, Japan and Brazil 27
4.2 Technology: growth of data facilitator companies and accompanying issues ... 27
4.3 Policies: role of public agencies in regulating big data in the transport sector .. 31
4.4 Externalities of big data technologies 34
4.5 Summary 36
5 National Level: Application of Big Data in the Governance Process 38
5.1 Big data governance by governments: general issues 38
5.1.1 Internal Governance: big data users 38
5.1.2 Applying big data in service providing and policy making 40
5.1.2.1 Service provision 40
5.1.2.2 Policy making 41
5.1.3 Challenges 42
5.2 Governance in transport 45
5.2.1 Big data in transport policy making and service provision 45
5.2.1.1 Transportation planning 45
5.2.1.2 Traffic monitoring 48

5.2.1.3 Public Transportation ... 48

5.2.2 Challenges ... 49

5.2.2.1 Making use of all the data ... 49

5.2.2.2 Limits of big data ... 50

5.2.2.3 Big data methods ... 50

5.3 Summary ... 50

6 International Level: Big Data in International Relations ... 52

6.1 How cross-border data flows are regulated ... 52

6.2 Motivations regulating cross-border flows ... 54

6.3 Effects of regulating cross-border flows ... 55

6.4 Summary ... 56

7 Privacy Issues of Big Data ... 57

7.1 General issues ... 57

7.1.1 The value of personal data ... 57

7.1.2 Why privacy is a concern ... 58

7.1.3 Big data and predictive analysis ... 59

7.1.4 Big data and security ... 60

7.1.5 Dealing with privacy concerns ... 60

7.2 Transport specific issues ... 61

7.2.1 Examples of various privacy issues ... 61

7.2.1.1 Vulnerability of transit passes ... 61

7.2.1.2 Privacy and ITS ... 62

7.2.1.3 Privacy vis-a-vis security ... 63

7.2.2 Dealing with privacy concerns ... 64

7.3 Summary ... 65

8 Big Data and the Centralisation of Market Power ... 66

8.1 Does big data lead to reduced consumer welfare? ... 66

8.1.1 Operational efficiency: scale effects, networking effects and high entry barriers 66

8.1.2 Loss of quality and innovation ... 67

8.1.3 Loss of privacy ... 68

8.1.4 Data Mergers ... 68

8.2 Summary 68
9 Conclusion 70
9.1 Approach 70
9.2 Analysis framework 70
9.3 Findings: per chapter 70
9.4 Findings: opportunities and challenges 72
References 74

List of Figures

Figure 1 Examples of the qualitative variety, volume and velocity of big data sources in travel and transportation companies (IBM 2014) 8
Figure 2 Depiction of overlap between big data business models and the transport sector 8
Figure 3 Underlying questions of D2.1, according to ToR 9
Figure 4 Visual depiction of the structure of D2.1 10
Figure 5 Penalties imposed by the GDPR 23
Figure 6 Country of origin for big data degree programs (EDISON, 2016) 30
Figure 7 Spending on ICT investment in 18 countries in OECD, Percentage, 2001-2010 31
Figure 8 Comparing the global electricity demand of the ICT sector with the illumination sector and electric vehicles (EV) sector. Source: (Mills 2013) 35
Figure 9 Four main components of the ICT ecosystem and relative share in power usage. Source (Mills 2013) 35
Figure 10 Big data governance framework for governments (Kim & Cho, 2017) 39
Figure 11 The policy cycle (Höchtl, Parycek, & Schölhammer, 2016) 41
Figure 12 Number of restrictions on cross-border data flows (vertical axis) between 1960 and 2017, based on 64 countries. Source: (Ferracane 2017) 53
Figure 13 Direct and indirect regulatory barriers to the free flow of data existing in Europe. Source: (Graux, Ypma, and Foley 2017) 54

List of Tables

Table 1 Data supplier companies and data user companies in EU for 2016 and 2017 26
Table 2 Data workers in the EU for 2017 26
Table 3 Big data Platform Vendors 28
Table 4 The application of BDA in the policy life cycle, based on (Höchtl, Parycek, and Schöllhammer 2016)....... 42
Table 5 The BDA framework for policy making applied to transportation, based on (Höchtl, Parycek, and Schöllhammer 2016)... 47
Table 6 Economic and political opportunities and challenges regarding the usage of big data identified in this report, divided into general and transport specific issues. 72

Glossary

Abbreviation	Expression
ACEA	European Automobile Manufacturers Association
BDA	Big Data Analytics
BDaaS	Big Data-as-a-Service
EDR	event data recorder
ENISA	European Agency for Network and Information Security
ERP	Electronic Road Pricing
ERTMS	European Railway Traffic Management System
EU	European Union
EV	Electric Vehicle
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
ICT	Information and Communication Technologies
IoT	Internet of Things
ISA	intelligent speed adaptation
ITS	Intelligent Transport Systems
LeMO	Leveraging Big Data for Managing Transport Operations
LRIT	Long-range identification and tracking
MOOCs	Massive Open Online Courses
ORION	On-Road Integrated Optimization and Navigation system
PC	Personal Computer
PDS	Public Data Services
PNR	Passenger Name Record

RFID	Radio Frequency Identification
RIS	River Information Services
SANRAL	South African roads agency
SESAR	Single European Sky Air Traffic Management Research
SME	Small- and Medium Enterprises
SSA	Shared Situational Awareness
SSN	SafeSeaNet
TEN-T	Trans-European Transport Network
ToR	Terms of Reference
TSC	Transport Service Customer
TSP	Transport Service Provider
WP2	Work Package 2

1 Introduction

The application of big data analytics in the transport sector is diverse – in terms of types of data, aims, functions, technology, and application area – and continues to develop with sophistication. These applications are carried out by individual actors or collectives, usually targeting small and well-defined functions. Nevertheless, the repercussions and effects – both positive and negative – can extend to the rest of the transport system, the economy and society it serves, and political entities. Taking stock of these direct and indirect effects is an important strategic step to designing effective policy and research roadmaps for the application of big data in the transport sector.

In this deliverable, the outcomes of Task 2.1 "Economic and political issues" of the Leveraging Big Data to Manage Transport Operations (LeMO) project are reported. The task involves reviewing the economic and political issues stemming from the existing and potential application of big data in the transport sector. Though the field is still developing, much has already been discussed in both academic and industry literature about the impact big data brings. The report briefly summarizes the main arguments and opinions found in existing research.

1.1 Abstract

The proliferation of data availability is reshaping the economic and political realm. On the one hand, big data enables private and public parties to provide better quality services and products. On the other hand, the exploitation of big data has led to policy response for limiting (e.g. the GDPR) or enabling (e.g. EU's policy towards a 'digital single market') the usage of big data.

This report aims at revealing the wider economic and political issues involved with utilizing big data by elaborating on the interaction between transport actors (demand-, supply-, external- and governance actors) and their role in the data economy (as data users, suppliers or facilitators). Subsequently, the interaction between these actors is described on various levels:

- Firm level
- Industry level
- National level
- International level

Having discussed this, the most critical challenges for the future identified in this report are:

- Lack of data professionals, in particular within governments;
- Government compartmentalization, limiting optimal data usage by governments;
- An insufficient framework that satisfies both the user demand for privacy and the usage of personal information for business innovation. This is particularly relevant for public-private data sharing;
- Too little awareness on the capacity of big data, as 'bad' big data analysis happens quickly, i.e. too little knowledge on what transport-related questions it can and cannot address, and how big data should address these questions.

1.2 Purpose of the document

The Leveraging Big Data to Manage Transport Operations (LeMO) project will contribute to formulating a research strategy for realizing a big data economy through considering the opportunities and challenges associated with big data in transport sector.

The collection, use, sharing and linking of transport big data implicates a number of economic, legal, social, ethical and political issues, including those which may result in positive and negative societal impacts. WP2 aims at mapping these wider societal issues in four separate deliverables. Together, the deliverables present a preliminary investigation of the wider societal issues that may be relevant to the consequences produced by big data, and examine how members of the public may perceive these potential consequences. This document, deliverable 2.1, focusses specifically on the economic and political issues related to big data.

Economic issues are identified by examining the *micro* (firm) level, *meso* (industry) level, and economic externalities of the data economy.

- On the level of the firm, specific transport related data innovations are considered.
- On the industry level, we map out the features of the European data economy by describing the supply-side (data suppliers and vendors) and demand-side (data users). In addition, the position of EU in the data economy as a whole is discussed as well as insight in the future direction.
- Lastly, externalities of the data economy are discussed, in particular the effects on the environment.

Political issues deal with the role of big data in the governance by governments. Two levels are discussed, the national and the international level.

- The national level addresses the question how governments, whether national or intergovernmental organisations such as the EU, can apply big data in its organisational structure, in policy making and in public service provision. Specifically for transport, big data in the government task of traffic monitoring, public service provision and transport planning is considered.
- The international level elaborates how borders impact the cross-border flow of data and why.

In addition, two major 'vertical' issues identified are examined in more detail: privacy issues and issues related to the concentration of market power. Both topics produce negative externalities with high societal consequences and require a more thorough investigation.

Altogether, the economic and political issues are redefined in general and transport related opportunities and challenges. In this way, LeMO can use the information of this deliverable as a springboard to further investigate research opportunities, challenges and limitations in relation to specific big data case studies.

1.3 Target audience

The results of this document are valuable for any private and public organisations that wish to inform themselves about the wider societal context in which transport related big data technologies are applied. This document deals specifically with the economic and political issues. Other deliverables under WP2 deal with legal issues (D2.2), ethical and social issues (D2.3) and rebound effects (D2.4).

This document does not provide insight in the specifics of how big data should be used, such as data managing technologies, which are described in detail in D1.3 on 'Big Data Methodologies, Tools and Infrastructure'. Rather, this document is valuable in the **objective definition stage** of big data applications in transport. It facilitates the discussion on *if* big data should be used by providing insight into *how* that may be achieved. Economic actors are provided with an overview of existing applications of big data in the transport sector. Political actors are given insight into how governments are currently using big data. The table of opportunities and challenges presented in the conclusion (see chapter 9.4) are helpful to determine which opportunities there are for applying big data or which challenges are to be overcome.

Additionally, the information in this deliverable can be used to evaluate current big data applications. Questions that policy makers or economic actors may ask could include: Are we dealing with all challenges that may arise? Are we exploiting every opportunity there is? Which challenges lay ahead in the future?

After reading this report and the other deliverables written under WP2, economic and political actors in transport are properly geared to determine if big data can be a solution to their question and in what way that may be.

2 Framework for Analysis

As the study covers a large domain, it would be good to define the boundaries of the study. An important starting point is to define the *stakeholders* of the application of big data in the transport sector. This is done by first discussing the stakeholders in transport (chapter 2.1) followed by the specific stakeholders in big data (chapter 2.2). Lastly, our understanding of the economic and political domain is explained in the dimensions of analysis (chapter 2.3).

2.1 Stakeholders in the transport sector

As Miles (2017) noted, there are many competing definitions of stakeholders. Using Miles' classification, it refers to entities, individual or collective, that have either one or a combination of the following conceptual roles:

- Influencer: one who "has the capacity to influence the actions of an organisation and has an active strategy to do so".
- Claimant: one who "has a claim on an organisation and an associated active strategy to pursue the claim but lacks the power to guarantee that the claim is attended to by management".
- Recipient: one who "is a passive recipient of the impact of organisational activity".
- Collaborator: one "that cooperates with an organisation but lacks an active interest to influence the organisation".

For this conceptual framework to have any practical use, it must be applied in well-defined contexts, as the roles undertaken by *real entities* can overlap and change. For example, a policy maker can be both a "recipient" of a transport service as a commuter or an "influencer" in the role of designing transport policy. Besides the action-perspective, the stakeholder will also have different motivations (i.e. disinterest, avoidance of risk, maximization of rewards, etc.) and different sense of responsibilities.

It is not the aim of this study to profile stakeholders in-depth, as the scope would be too demanding for the entire transport sector and data economy. Instead, we will note down the common types of stakeholders – in terms of their functions in the transport industry and in the effects of transport externalities. A systematic way of grouping the stakeholders is to look at the following aspects of transport:

1. Transport demand
2. Transport supply
3. Transport externalities
4. Transport governance

2.1.1 Transport demand

Transport demand is commonly considered a "derived demand", where demand for transport services emerge as a result of activities (requiring people and commodities) carried out in different locations. Hence, demand for vehicles, transport services and transport infrastructure is strongly influenced by the spatial structure of residential, commercial, industrial and administrative areas. More importantly "effective transport demand", in contrast to generic need, is characterized by conditions set by the potential transport service customer, in terms of willingness-to-pay (in price or other units) and quality of service.

At a very high-level, transport demand can be segmented by transport object (persons and goods) and by spatial-region (e.g. rural, urban, regional or international). The role of marketing is important to define clearly the boundaries of more segments. Passenger transport is commonly divided into business, commuting or leisure trips. Goods transport is more heterogeneous and can be divided into categories depending on product (e.g. refrigerated or bulk goods) or supply chain (e.g. distribution or reverse logistics), specific services (e.g. courier, postal services or home delivery), or by spatial characteristics (e.g. long haul, domestic shipping, or last mile). Goods transport is particularly complex, since this sector has a great diversity of goods and the corresponding optimal conditions for transport. Furthermore, there are intermediate layers, such as the logistics centres and intermodal hubs that further segment the transport demand in separate trip chains, which behave very differently.

2.1.2 Transport supply

Transport supply is the activity that satisfies transport demand. Both the transport demand and supply characterize the transport service market. The categories developed in the SMARTFREIGHT[1] project (Natvig 2009) are useful to identify the roles in the transport market. These roles may sometimes be carried out by the same actors. These are briefly described below:

- **Transport user** has a transport demand, i.e. a need for a passenger or freight to be transported, or organizes the transport operations on behalf of others. The transport user is a customer to the transport service provider and makes a request, specifying the type of transport, the details of the transport and quality required, and agrees to the price stipulated by the transport service provider.
- **Transport service provider** has the central role of providing the transport service to the transport user, within the constraints and limits of vehicle resources and the transportation network it uses.
- **Vehicle operator** carries out the transport activity for the transport user. It works within the framework of the transport service provider, who sets the terms of the service.
- **Transportation network manager** is responsible for building and maintaining the transportation network infrastructure. They rely on forecasting and planning for demand to ensure a particular level of service. The transport network refer to both the "links" and the "nodes". Though links are often the more visible form of infrastructure, nodes provide a necessary service to facilitate transfers from transport mode to another, and to provide other logistical functions.
- **Transportation network resource manager** is responsible for managing the utilization of the transportation network. These are the operational decisions that need to be made, taking into account the capacity and capability afforded by the built infrastructure and the level of traffic demand that emerges from the transport service providers. Note also that

[1] The SMARTFREIGHT project is founded by EU's FP7 and aims at improving freight distribution in urban areas

efficiency is not always the only criteria to be optimized. Recent years have emphasized the need for sustainability, security, and even system-wide criteria.

Other actors that are relevant are the actors in the supply chains supporting these transport service actors, such as **vehicle manufacturers** who provide transport vehicles. As the vehicle and transport systems are increasingly dependent on software and digital infrastructures, the prominence of **software engineering firms or digital service providers** may be considered.

2.1.3 Transport externalities

Each voluntary actor in an economic exchange in the transport market profits from the exchange – e.g. the user receives the service while the transport service provider receives an income. There are effects that occur outside of this exchange, which are termed **transport externalities**. 'Externalities' is an economic term, which refers to "connections, relations and effects which agents do not take into account in their calculations when entering into a market transaction" (Callon 1998).

While it is often associated with negative externalities, some effects may also be positive (i.e. bring a net benefit to actors outside the transaction). The more common categories of negative transport externalities are air and noise pollution, emissions of greenhouse gases, accidents, dependence on oil (and its price and supply volatility), congestion, track damage (i.e. road and rail damage), and social exclusion (Santos et al. 2010). Some positive transport externalities are the increase in jobs in a particular area and agglomeration effects.

The existence of negative transport externalities gives rise to "recipient stakeholders" (see section 2.1). These roles are often undertaken by **citizens or residents**, represented by **activist organizations** or **public agencies**. Public agencies may also have "claimant" or "influencer" stakeholder roles, depending on how they intend to influence the mitigation of transport externalities. As transport externalities are often thought to be a symptom of "market failure", government agencies claim that measures must be taken to correct these failures (Santos et al. 2010).

2.1.4 Transport governance

Transport is one of the most regulated economic activities for two main reasons. Firstly, it is thought that government oversight is necessary to mitigate negative transport externalities. Various public agencies and policies have been designed to ensure that negative impacts, such as environmental damage are mitigated. Regulations and incentives are therefore introduced to reduce individual vehicle pollution, reduce transport activity in general or locally, encourage better transport modes. These agencies operate at the local to national level, and beyond.

The second reason that transport is heavily regulated is because of its strategic function to enable trade, drive the economy and to maintain mobility of the residents. Hence, much transport infrastructure is built, maintained and operated by the government, if not at least subsidized and strongly influenced by government. Urban public transport also often does not operate without significant government intervention. Transport policy is aimed at mitigating risks of failure, improving connections and driving growth of regions or economic sectors.

Hence, under transport governance various public agencies with different portfolios (not only transport, but also land use, environment, energy, economy, etc.) are tasked with managing parts of activity, effects or resources in the transport industry.

2.2 Role of big data in the transport sector

The premise of the study is that big data can play a role in the activity carried out by each stakeholder in the transport sector; from user to service provider, from resident or activist to public agency. To understand their role, we can first look to generic business models in the big data economy, each of which can be undertaken by any of the transport stakeholders discussed in the preceding chapter.

Three high-level big data business model categories were proposed by Schroeder (2016):

- **Data users** are organizations that "use data to inform business decisions, or as an input into other products and services such as credit reports or targeted advertising campaigns" (Schroeder 2016). Data users **source, analyse and link data** to arrive at their insights. The aim of data users fall under three main categories: product innovation, process innovation, and sustainability measures.
- **Data suppliers** are organizations "that either generate data that is of intrinsic value and therefore marketable, or else serve a kind of brokerage role by providing access to an aggregation of first- and third-party data" (Schroeder 2016). The data generated may be a side-effect of digitalization of transactions or management tools, or perhaps even the actual focus of data collection, such as the management of traffic sensors. Data brokers are a subset, who focus on collecting data from various sources and providing it to others for a fee. Data suppliers may also be from entirely different fields, unrelated to transport, such as social media networks. Data suppliers **produce and share/sell data**. The production of data depend on the policies and contracts governing data subjects. Terms restricting the sharing and selling of data may be included in the policies and contracts.
- **Data facilitators** are organizations that "perform a range of services including advice on how to capitalise on big data, the provision of physical infrastructure and the provision of outsourced analytics services" (Schroeder 2016). Data users and suppliers outsource their functions either partially or completely to them, because they either "lack the expertise or capacity" to do so themselves, for cost or strategic reasons. Since they are a third-party provider, their roles and responsibilities are primarily defined by the contractual terms with their customers - data users and suppliers.

For the most part, the main role played by the transport sector actors are as data users and data suppliers. As big data may depend on a wide variety of data, it is important to ensure that their data suppliers' role is not ignored (see Figure 1). This might involve public sector agencies, social media companies, advertising agency data, telecommunications companies, IoT services, etc. Data users will have to be creative in identifying the benefits (and discerning the costs) of obtaining these different data sources.

Supporting the data users and suppliers are data facilitators who are primarily in the analytics industry, ICT, or database systems industry. While there may be products that focus on the transport sector, these are rare and few.

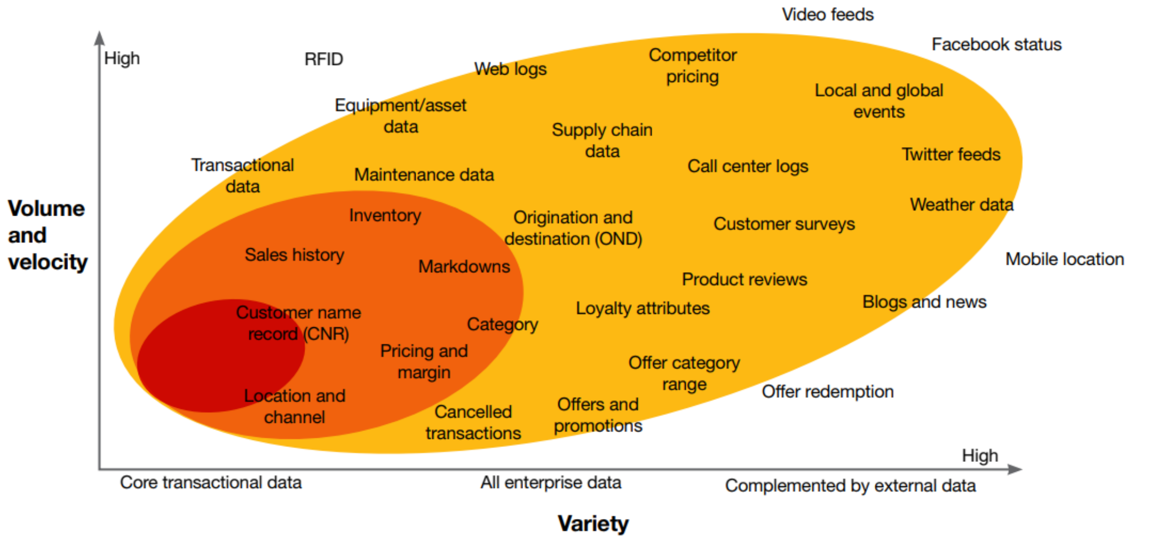

Figure 1. Travel and transportation companies must consider an increasing amount of information to make strategic decisions.

Figure 1 Examples of the qualitative variety, volume and velocity of big data sources in travel and transportation companies (IBM 2014)

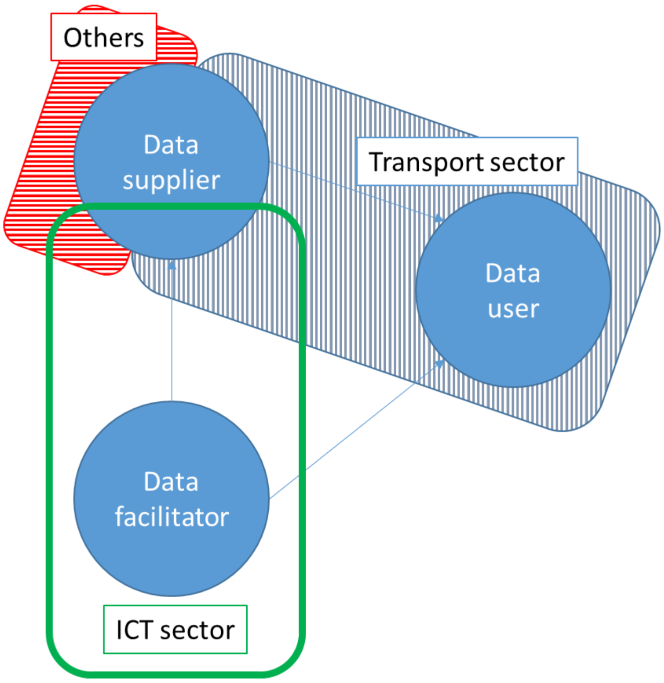

Figure 2 Depiction of overlap between big data business models and the transport sector

Figure 2 shows the overlap of the big data business models with the transport sector. Our understanding is that the transport sector will include both data users and data suppliers, but rarely if ever any part of the data facilitators. The useful data for the transport sector will mostly come from within data suppliers within the transport industry, but also potentially from the outside, such as the public sector, weather agencies, or academic institutions.

The public sector may also play a role in regulating the flow of data, in addition to their role in providing data – whether open or for a fee. As we have examined in D1.2 on 'Big Data Policies', the main interests (so far) in the regulation of data are protecting privacy, and encouraging open data, digitalization and optimization of the transport sector.

2.3 Dimensions of analysis

The underlying questions regarding D2.1 and their relationships are described conceptually below (*Figure 3*). The figure shows the way in which the actors described in chapter 2.1 and 2.2 interact with each other.

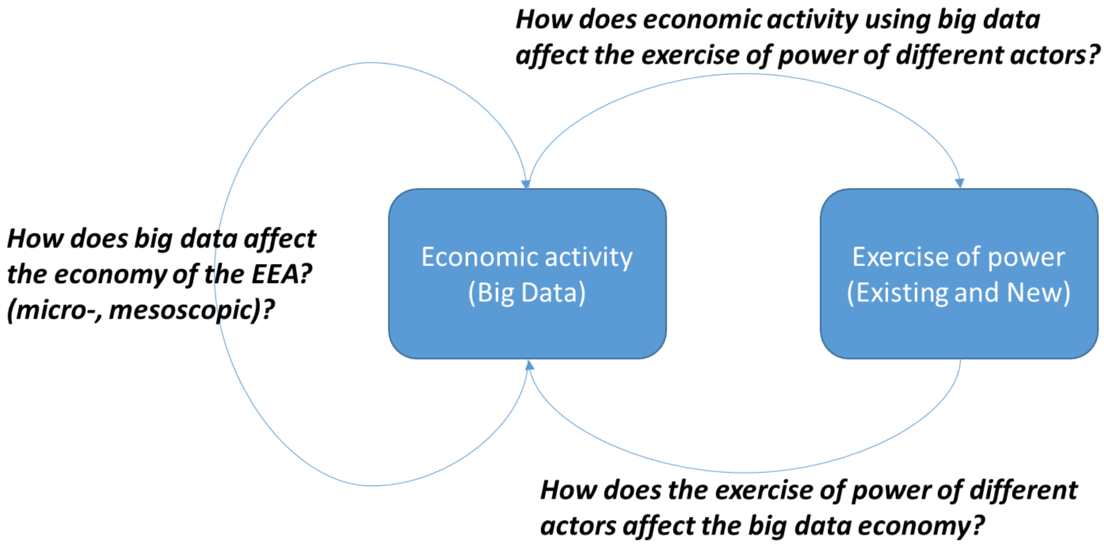

Figure 3 Underlying questions of D2.1, according to ToR

These questions are addressed by describing the ways in which various actors (transport demand, supply and governmental actors) use big data as well as the consequences of its usage for other actors in society (private and public). The following structure is used to address these questions (*Figure 4*, below).

2.3.1 Economic dimensions

From an economic point of view, the influence of big data, whether in the use of big data analytics (BDA) in the transport sector or the development of the data market, can be viewed at two main levels at the firm's level (micro) and at the industry level (meso). Examining the influence at the firm's level is easy to understand, since firms will implement and take part in the economic activities using or supplying big data. This primarily focuses on whether big data innovation will improve the market offering, improve how the services can be provided, and reduce externalities. This will be discussed in Chapter 3.

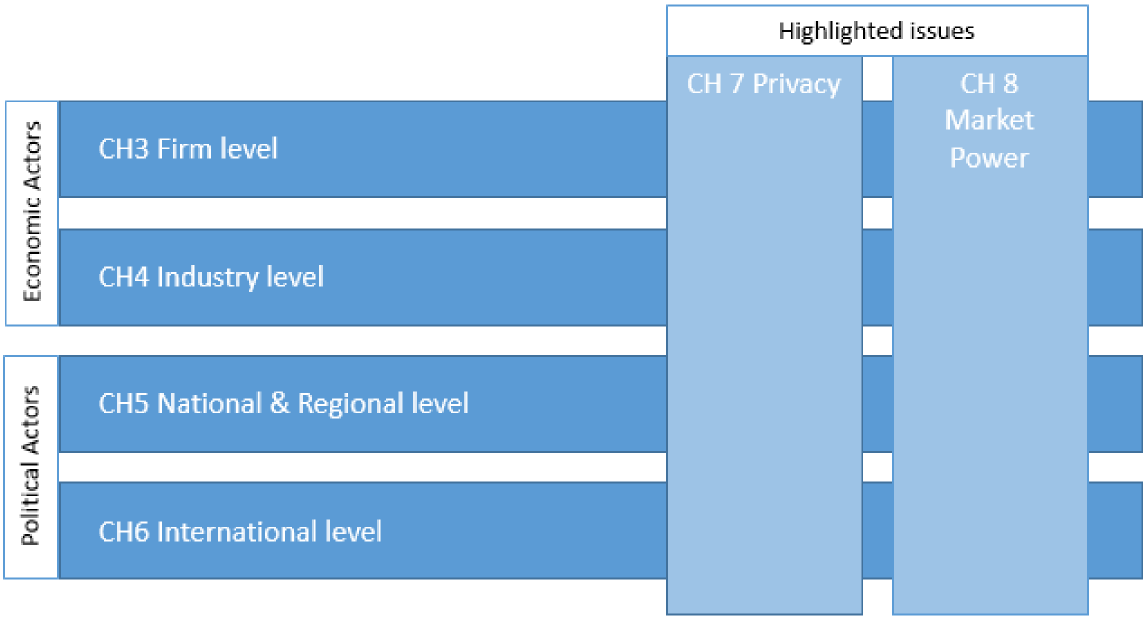

Figure 4 Visual depiction of the structure of D2.1

At the industry level, we can examine how the industry and market will evolve when innovations or regulations are implemented by one or more firms. Firms that successfully innovate can acquire a leading role and cause a shift in market power and competition levels that affects the rest of the industry. An important question is whether valuable data (for a particular BDA application) or technology will be held by one or only a handful of firms, thus reducing contestability of the industry – a key ingredient in a competitive market. There are also other aspects that need to be considered at the industry level, which will be covered in Chapter 4.

2.3.2 Political dimensions

From a political point of view, the goal of using big data is to exploit the positive externalities of big data while limiting its negative externalities. The role of the governments is to regulate and enable big data in such a way that it contributes to the well-being of society.

Chapter 5 elaborates on the application of big data by public parties in the way they govern society. Discussed are how governments use big data in their own organisational structure, in policymaking and in public service provision. Another major role of governments is making big data policies. Current big data policies are already discussed at length in D1.2 on 'Big Data Policies' and will not be covered here. With regards to transport governance, the following tasks of the government are considered: transport planning, transport monitoring and public transport provision. Chapter 5 ends with a summary on the challenges and opportunities related to governments applying big data. Based on the opportunities and challenges big data presents, future policy topics are identified.

When data crosses the border it becomes a matter of international relations. The impact of borders on (transportation) data flows are discussed in chapter 6. For example, governments can decide to apply restrictions or conditions on the storage or processing location of certain kinds of data. On the one hand, limiting the movement of data gives governments more control over data. At the same time, organisations can be limited in their innovations when data cannot be fully exploited due to movement restrictions. The underlying motivations for nations to apply regulations to cross-border data transfer are considered to identify the main challenges and opportunities.

2.3.3 Highlighted issues: privacy and market power

Having explored the impact of big data to national and international politics, in the next chapters two major big data issues are examined in more detail: privacy issues and issues related to the concentration of market power as a result of exploiting big data. Both topics produce negative externalities with high societal consequences and require more thorough investigation.

Chapter 7 deals with the issue that the spread of personal data endangers people's privacy. If the general public becomes reluctant to share their data, it may be detrimental for innovation. Transport policy and transport services heavily rely on location-based data, and the absence such data limits the quality of service that transport companies can provide. Chapter 7 highlights the privacy issues and presents potential solutions.

In Chapter 8 the issue of data power held by private parties is discussed. Recently, a few powerful and massive companies have emerged whose growth is based on the exploitation of big data. The presence of a few big players holding a large market share has brought new challenges to anti-trust authorities related to harm to competition and consumers.

2.4 Summary

In the section, the framework of understanding the economic and political issues of big data in the transport sector have been introduced. A fundamental aspect to discuss are the different types of actors in the transport and big data economy. In the transport sector the actors are examined by considering four aspects of the transport economy: demand, supply, externalities and governance. In the big data economy, the actors are examined from a simple value chain perspective: data users, suppliers and facilitators.

The next step was to identify how the actors interact in the political and economic realm. The relationship of the various actors are considered according to the following questions:

- How does big data affect the economy on a micro and meso level, and for the EEA in particular?
- How does economic activity using big data affect the exercise of power of different actors?
- How does the exercise of power of different actors affect the big data economy?

In the following chapters, the issues are discussed in further detail. Economic issues are discussed in chapter 3 (firm/micro level) and 4 (industry/meso level). Political issues in chapter 5 (national level) and 6 (international level). The cross-domain issues of privacy and market power centralisation is explained in chapter 7 and 8.

3 Firm Level: Influence of Big Data on Business Models of Transport Companies

In this section, we consider the influences that using or supplying big data will have on transport sector actors. While transport services can be expected to be directly improved through BDA, its effectiveness and costs will vary depending on the industry and of course the direct implementation. We will also consider the potential role that transport sector actors could play as suppliers of data, which is an important element in developing a data economy for and within the transport industry. Finally, we will discuss several aspects of costs for a company interested in moving towards using or supplying big data.

3.1 Direct improvements to the transport services

3.1.1 Shared situational awareness of the transport network

As we have previously described, big data analytics (BDA) enables description of the transport system. A basic, but practically challenging skill that transport infrastructure managers need is "situational awareness" of the network they are responsible for managing. Shared Situational Awareness (SSA) is defined as a "shared awareness/understanding of a particular situation" or "common operational picture" about a problem situation (Nofi 2000). Network managers aim for this in different ways according to the types of transport network they are responsible for and the types of data they can collect.

While many systems are still primitive, the increasing digitalization, the use of Internet of Things (IoT), the improvement in real-time communication, i.e. in short, the availability of big data accessible to the transport network managers, have also given rise to BDA to provide SSA. SSA can be provided simply by ensuring the right types of data and insights derived from the data are available to the relevant parties. Actions can then be decided and acted by the individual party. The benefits of SSA of the transport network is the transport network managers to have the necessary information to manage delays, failures and disruptions, adjust transport supply to meet variations in transport demand, support safety of transport users, and support customers in planning their own transport activity. SSA is also beneficial in managing supply chains, which drives demand for freight transport.

3.1.2 Improving transport network capacity for all modes

The primary aim of transport network managers is to provide network capacity that satisfies the transport service providers' requirement for quality and price. There are several key functions that are necessary: (1) long term planning of the transport network, (2) real-time traffic management, and (3) mitigating risks that degrade network capacity. These issues are applicable to all modes ranging from road to rail, water to air. Here, we present several examples of research and applications for the three functions.

3.1.2.1 Long term planning of the transport network

As previously discussed, transport activity is a derived demand. This offers a basis for transport vehicle demand to be predicted in the long term, depending on the forecasted demand for passenger and goods movement. While some segments of the transport network must be provided to support the "right of passage" or accessibility to residences and services, the design of the transport network is mostly driven by models predicting transport demand.

Transport network design models can incorporate a large amount of data, with the purpose of increasing its accuracy in predicting transport demand and fulfilling it. This demand can be derived from deriving traffic observations from the existing transport network and combined with forecasts of future population and economic development (Milne and Watling 2018). Big data can also be used in conjunction with urban planning (Glaeser et al. 2018; Batty 2013) a public policy and design aspect that often goes hand in hand with design of the transport network.

Big data can also enable a better understanding of existing supply issues and bottlenecks in the transport network (part of the SSA). The use of sensors, ticketing data and other similar data streams can be used to infer more precisely how the transport network was used (in time and space) and by whom (customer/transport user), rather than relying, as previously on macro-indicators (such as ticketing volume, traffic volume, etc). Collecting these insights in the short term can be used to give a long term appraisal of the transport network performance, which supports the planning.

3.1.2.2 Real-time traffic management

The operational aspect of transport network management requires in many instances real-time traffic management. A simple and ubiquitous instance of which is the traffic signals at a road intersection. Traffic signals, though previously were done by hand, have been automated, unless in extremely disruptive situations. Traffic control systems include intersection control, ramp metering, dynamic highway lanes, public transport prioritization, route diversion messages, etc. Still, though automated, the real-time traffic management *are not dynamically adjusting to real-time transport demand*, but rather rely on a fixed set of predefined plans that are triggered based on the conditions of the road or in time.

The use of big data to enable these systems to react to changes in transport demand and allocate transport capacity at different links is still being researched, but holds much potential (Zhang et al. 2011). These are so-called "data-driven" intelligent transport systems (ITS), as opposed to "model-based" ITS. The main benefit is that the data used to underpin control systems are based on **what is** (to the degree the data is precise and accurate), rather than on what a theoretical model (and therefore inherently *idealized*) proposes is happening. Big data architectures can also be used to account for other aspects of the transport system, linking for instance, public transport service providers, rail networks, events management systems, and incident management systems. All of these systems have a bearing on the need for transport capacity in the road network and can potentially be included in the traffic control systems.

Besides traffic control systems, a function of growing importance is road pricing. Road pricing, under which congestion charging also falls under, is a method to internalize the costs of **using a particular segment of the road network when transport capacity is scarce**. (This differs from road taxes, which does not influence transport demand in real time or fixed toll-roads that are always charged to the user regardless of time.) While London implements congestion charging, a well-discussed case is the implementation of Electronic Road Pricing (ERP) in Singapore. In its current form, motorists are charged rates according to the gantry passed, the time of the day, and the type of vehicle. The rates are published quarterly and are therefore **not dynamically** adjusting to the real-time available transport capacity, but rather to rates defined by historical data. A proposed system to be implemented in 2020 will use GPS-coordinates and will charge based on distance, instead of entry. This might introduce an element of reaction to transport usage, thus increasing the disutility of private vehicle travel. However, it is unknown whether the rates will actually depend on **nowcasted** congestion or be a fixed rate based on historical data of congestion.

A more proper congestion charging based on BDA is being implemented on a highway between Jerusalem and Tel Aviv (Siemens 2012). The toll system gives road users the opportunity to choose for the fast lane with a toll or the slow lane. The toll price adjusts based on the dynamic traffic volume (an indicator for congestion). This adjustment allows users to pay to reduce their travel time, when it is valuable (i.e. heading to work in the morning) or to accept a longer travel time, when it is not urgent (i.e. return from work in the evening). Users can also consider park and ride systems in real time, hence focusing mobility, rather than vehicle throughput.

Similar potential use of BDA for managing traffic capacity can be found in rail (Ghofrani et al. 2018), maritime (Zaman et al. 2017; Lind et al. 2016) and air traffic control (Belcastro et al. 2016; Keller et al. 2016). The key aim here is that transport network managers need to have the tools to reduce vehicle traffic in general or redirect traffic to less used network links. BDA can support this function, however the limitations of the effectiveness is the control method used. An open question is whether the transport users are sufficiently flexible to be redirected and what mechanisms (such as road pricing) can big data additionally support.

3.1.2.3 Mitigating risks that degrade network capacity

Besides reacting to the impact that transport demand reduces transport supply (as congestion does), transport network managers must also defend itself from other events that are damaging both in the short term (such as disruptions to the transport flow or safety) or in the long term to the network assets.

Traffic incident management, arguably a subset of real-time traffic management, can also be enhanced through BDA for monitoring and estimating causes of incidents (Steenbruggen et al. 2011). This deals with **unpredictable congestion**, such as bad weather and traffic incidents. Incident management systems need to be quick its response and provide sufficiently competent mitigating measures, to return the traffic system to “normal”. Prescriptive BDA can be applied to support quick decision making.

A special case can include events organized (such as concerts or sporting events) that have a significant momentary impacts on the transport system. In this case, transport network and service providers need sufficient data to allocate network and service capacity. Ideally, this will involve multi-modal services to deal with both private cars, non-motorized modes, and public transport services (Papacharalampous et al. 2015).

Mitigating **failures in network assets** is an important function of transport network managers. For railway operators, BDA can support more cost-efficient and accurate maintenance strategies. Literature has identified three strategies for maintenance: **corrective, preventive and condition-based maintenance strategies**, which can be directed to *maintaining* the vehicle, track and signalling equipment (Ghofrani et al. 2018). Since corrective maintenance occurs "after a defect or failure occurs", preventive and condition-based maintenance strategies are preferred. Preventive maintenance involves scheduling of "adjustments, major overhauls, replacements, renewals, and inspections" to reduce failure of equipment during operation. Condition-based maintenance involves predicting the best schedule for maintenance activities by monitoring the conditions of individual sub-systems or components in the system. Both strategies can strongly benefit from the use of BDA in predicting failure (Jamshidi et al. 2017; Li et al. 2014), if the data is available.

3.1.3 Improvement of transport services

Assuming good quality provision of transport network services, transport serviced providers (TSP) have the most influence on whether the transport service customer (TSC) is satisfied with the quality and price of the service – the two elements crucial in the purchase of the service. We have identified several ways that BDA can support the TSPs to improve their services (in quality and price).

3.1.3.1 Marketing activities in the transport sector

Hibbs expresses marketing in the following way "in transport, marketing means making sure that what <u>we produce we can sell</u> and that <u>we sell what we produce</u>, <u>at a price that satisfies both user and seller</u>" (Hibbs 2003). It is clear that for a transport company to thrive or survive, they need to be sufficiently competent in *marketing*.

There are four practices in marketing that are vital to understand: **market intelligence, marketing research, marketing communications and promotion**.

- Market intelligence aims <u>to understand the market</u>, specifically the target market (both the demands of the customer and the willingness and capability to pay), the actual performance and capability of the firm to meet the demand, and the performance of competition (Hibbs 2003). In the competition between public transport services and private cars, understanding the major benefits and disbenefits of both modes is a major step in designing public transport services to *better* fit the customer's demand for mobility (Beiraõ and Cabral 2007).
- Marketing research aims <u>to understand and plan the firm's role in fulfilling the customer segment's demand requirements</u>. It would build partially on market intelligence, but would look at ways how the firm could fulfil demand. For instance, understanding the determinants of selecting freight carriers can support the firm in its investment planning and promotion activities (Solakivi and Ojala 2017).
- Marketing communications, under which advertising falls, aims communicate to existing and prospective customers about the offerings of the firm. Targeted advertising, for instance by e-commerce merchants or tourist agencies, is a crucial aspect of the business (Akter and Wamba 2016; Minghetti and Buhalis 2010).
- Promotion (campaigns) aims to attract customers leading to greater sales. In promotion campaigns, the service offering is modified to provide better value to customers and is

then *promoted* to customers. An example are promotion campaigns to flatten public transport peak hour demand in Hong Kong and Singapore (Halvorsen 2015)

In each of the activities, BDA could play a role (Erevelles, Fukawa, and Swayne 2016), although specifically in the transport sector, research has not *yet* witnessed it. However, some examples of BDA on geospatial data “to predict location of a wide variety of hundreds of subjects even years into the future and with high accuracy” (Sadilek and Krumm n.d.) will certainly support marketing of new services to these areas (Khan, Ngo, et al. 2017).

3.1.3.2 Vehicle routing

From the vehicle operator’s perspective, routes from origin to destination should reduce (generalized) travel cost. Visualization of congested areas and calculation of potential delays is already being provided by companies such as Inrix and HereMaps. This is also integrated into navigation systems, such as Waze or TOMTOM.

Vehicle routing can be planned in advance, such as in the freight distribution sector application, where the sequence of stops for each vehicle must be planned, in order to identify sequence for loading of shipments in the cargo hold. Improving the process for route optimization can save a freight carrier a lot of money and significantly reduce environmental footprint (Parcu, Brennan, and Glass 2018). An example of real-time optimization is the On-Road Integrated Optimization and Navigation system (ORION) at UPS, which provides the driver with routes taking into account a large set of historical data, as well as up to date congestion and other relevant data[2&3].

3.1.3.3 Service provision and schedules

Transport service providers need to set the terms of the services, i.e. the specific service offerings, based on the firm’s value propositions. This is a necessary function of proper marketing research. As previously highlighted, BDA can be applied to understand demand for transport in a transport network.

The same techniques with more precision can be used to design a transport services network, for instance, in public transport network planning. Some of the design parameters are network topology, stop positioning, service design, and terminal and link capacities. To understand the amount of services that need to be provided (in time and space), we also need to understand passenger mobility demand (in time and space). Demand is commonly estimated using choice modelling or equilibrium assignment methods. But, there are several issues with *classical methods,* such as dealing with modifications of passenger routes when faced service delays, dynamic nature of traffic, day-to-day variations in demand (Ortúzar and Willumsen 2011).

BDA can support public transport network and service provision planning by utilizing the very rich electronic fare card data, which are personalized in many cities (e.g. Singapore, Seoul, and

[2] https://www.pressroom.ups.com/pressroom/ContentDetailsViewer.page?ConceptType=Factsheets&id=1426321616277-282,

[3] https://www.bsr.org/en/our-insights/case-study-view/center-for-technology-and-sustainability-orion-technology-ups

London) (Ma et al. 2013) and mobile phone GPS data (Anda, Erath, and Fourie 2017). This includes observations on disruptions to the schedules of public transport services, which can be used to identify and correct systemic issues in the network (Furth et al. 2003).

Research also shows how personalized data can be "anonymized" algorithmically, while preserving the richness of the trajectory, which can be useful to preserve the privacy of transport user (Ghasemzadeh et al. 2014). On the other hand, by preserving the identity of transport fare card holders, the service can be further personalized – thereby further segmenting the market, leading to *better* service provision. This was demonstrated with London's Oyster card data (with consent of the fare card holders) to test the potential of personalized travel time estimates, a significant factor in the provision of travel route recommendations (Lathia et al. 2013).

This similarly applies to the use in other transport service sectors. In real-time, transport supply can be increased in areas of high demand to accommodate the increase in demand. This requires SSA of demand, which can be gained through BDA. Of course, this also means that a control method should be applied that can work in real-time. For instance, "surge pricing" can *attract* drivers using the Uber and Lyft platform to areas where there are too few drivers for the number of passenger demands (Cachon, Daniels, and Lobel 2017). Though somewhat controversial as passengers are unused to prices increasing for taxi services, it is shown to effectively motivate Uber and Lyft drivers to provide services. Availability of taxi services in congested periods, risky moments and sparse areas have been a difficult problem to solve for taxi companies[4], since the drivers are paid per customer trip not at a fixed rate.

This also solves another economic problem in transportation, which is the pricing of services. Transport service providers have a challenge to set a price for their market services, which optimizes their own resources and costs, as well increases total revenue. An important concept here is that prices should reflect "what the market can bear" (Hibbs 2003). Any higher, the service provider loses customers; any lower, the service provider is not optimizing its profit. With a surge pricing policy, however, one can readily calculate and test exactly the prices that the market can bear. This can also inform other transport service providers on how to price similar services at the market prices. (Note that this is different from pricing discounts for special social groups, such as students or senior citizens).

3.1.4 Shift towards sustainable mobility and transport

We have discussed many ways in which big data can help in reducing waste and increasing efficiency, while maintaining quality of service and pricing. Further, we have considered how BDA can improve the planning, operation and marketing of public transport services. Both of which are the usual ways to increase sustainability in the transport sector. For instance, improving vehicle routing can bring immediate fuel savings to freight and passenger transport vehicle operators, which reduces air pollution and emissions of greenhouse gases. BDA was also shown to be useful in road pricing, which is a way internalizing the external cost of traffic

---

[4] https://www.bloomberg.com/news/articles/2013-09-19/taxis-vanish-in-rain-as-singapore-gets-congested-southeast-asia

congestion. Reducing travel, especially in peak periods, reduces the economic and social cost of congestion.

Here, we also propose another way that BDA aids in our quest for sustainability, evaluation of transport policy and solutions. Zolli (2018) discussed the idea of moving "from 'big data' to 'big indicators' to 'big instruments'", which simply means using big data to calculate social and environmental indicators that can help us understand the larger systems in more comprehensive ways, without being overloaded with information that are less valuable. Though there are indicators that are an important means to understand sustainability in transport, these are usually couched within very limited analysis frameworks (Cottrill and Derrible 2014).

Zolli suggested that the *understanding* is further automated and monitored in real time, and that ***big indicators*** are used rather than conventional indicators. They differ from conventional indicators, in terms of (Andrew Zolli 2018):

- **Precision and resolution**: Because they are based on vastly more real-time information than a typical indicator, a big indicator will tell us what is happening in a system with much greater spatial and temporal resolution, precision, and sensitivity. It might tell us about the growth of a city, a particular neighbourhood, and a given block with equal precision.
- **Frequency**: Since they are continuously recalculated, big indicators will tell us how a system is doing day to day, not just year over year. As such, they can help us answer questions like: Is a particular intervention working? Making things better or worse? Having intended or unintended consequences?
- **Scale and Reach**: Because big indicators take advantage of the global reach of new sensors and datasets, they can capture the state of systems previously too large, too remote, or too expensive to monitor effectively.
- **Predictive Capacity**: As we collect more and more data about the world, in many areas we are beginning to find not only the markers of past change, but also the early warning signs of impending change. Big indicators will not only tell us where the system is at a given moment, but also where it is likely to go in the future.
- **Sophistication**: As noted, many of the world's critical systems behave in complex and counter-intuitive ways. They also influence each other in subtle and unobvious ways. Big indicators will begin to illuminate these behaviours and connections so that we can create more effective, comprehensive, and long-lasting interventions.
- **Interoperability**: Big indicators will allow us to compare patterns and processes across jurisdictions. They will, in effect, become new measurement and reporting standards.

Rich, varied and real-time big data can be used to understand sustainability issues in transport, even if not presented as a *big indicator*. For instance, based on predictions of travel demand, we can evaluate the amount of traffic impacts that the current transport system causes. There are also already technologies in place to measure and collect this information, although not necessarily integrated from cause-and-effect. For instance, the Noise Observation & Information Service for Europe (http://noise.eea.europa.eu/) provides monitoring of noise levels, but these need to be *linked* to the sources of noise (such as road traffic) for proper policy making. Another example is the use of big data to understand the transport market better, particularly in terms of poverty or social exclusion (Njuguna and Mcsharry 2017). This can be used in transport network and service providers in supporting affordable public transport services.

Further discussion of how big data can be used in sustainability evaluation are presented (Cottrill and Derrible 2014).

3.2 Role as a data supplier

In the transport sector, firms (and public agencies) involved with providing network infrastructure and services may also play a role as data suppliers. What are the motivations for this?

- They may sell data for profit. For instance, it is alleged that Transport for London is selling data on passengers using the London Underground by tracking their WiFi receivers on their smartphones[5]. The data is collected primarily for the operator to "better understand journey patterns and improve [their] services". But, as a secondary revenue source, they are allegedly posed to selling it to third parties for a sum of £322m. Though the story concerns a case in 2016, the potential for creating a business model that sells data is still there.
- They may exchange data for other data. As documented in Section 3.3 of Deliverable 1.2 on big data policies, Waze's Connected Citizens Program collects and shares location data of its users to support "cities, departments of transportation and first responders" in planning and operations. In exchange, the departments of transportation shares real-time traffic and incidents information with Waze to improve its own services, helping to re-route drivers and manage disruptions.
- The data can be provided to third-party service providers who can use the data for enhancing the data supplier's own main business model. This is one of the motivations behind the open data by private sector firms also discussed in Section 3.2 of Deliverable 1.2. The airport for instance can provide open data for others to integrate their own services in it, such as to coordinate last mile passenger transport. Many digital transport network management systems also require data from its actors to increase its value to its partners.
- The data can be supplied to academia and researchers to discover new value. This is more of an explorative move to see what value specific data sets can create, if academia and researchers are provided with the data. This may or may not have any benefit to the firm.
- Data can be provided to promote the "common good". While the actual motivations of promoting the common good may differ, firms can decide to share their data to support "common good" causes. The applications are varied (as we have previously discussed in Section 3.1).

[5] https://news.sky.com/story/tfl-may-make-322m-by-selling-on-data-from-passengers-mobiles-via-tube-wifi-11056118

- Data sharing can be legally obligated. For instance, in air travel, regulations require passenger information to be shared with immigration offices in advance. In other cases, firms can be required to share their customs information for importing and exporting of goods.

It is important to note that transport sector actors hold an important role as data suppliers, but their main customers would be other transport sector actors. Hence, in many cases their customers are their "competitors", which would be a strong disincentive to share with them for free.

A study on data sharing between companies in Europe (everis Benelux 2018) surveyed companies to develop a general profile of data suppliers. The findings of their survey pointed out that data suppliers are generally large companies, sharing data within their own business sector. Their main motivation to share data is to develop new business models, services and products.

3.3 Costs of innovation

Innovating a business model may hold much potential, however there are other risks and costs associated with moving towards a big data oriented business model. Here we highlight some of the issues.

3.3.1 Bad or dark data

The fundamental resource in our analysis is the data produced, collected, measured, shared, or sold. However, there are two types of issues related to the data.

Bad data. An important reason why newer big data definitions consider the fourth V to be Veracity or Verification (Zhong et al. 2016) is because of the potential of bad data (e.g. noises, inaccuracies, delayed responses, corrupted data, etc.) to be mixed up with the *good data*, and how that may be catastrophic for BDA. Hence, there is a costly risk for both bad outcomes of BDA and the mitigation costs (e.g. verification processes). Also note that as a data supplier, the firm will also need to ensure that the quality of the data is high, which can also be costly if the bad data is not adequately detected.

Dark data. Dark data can be defined as "the information assets organizations collect, process and store during regular business activities, but generally fail to use for other purposes"[6]. In other words, there is data that the firm has access to (though not necessarily *owns*), which is not exploited to create value. It may include data that the firm cannot create value from, such as if the firm only needs real-time data, historical data is useless. The danger dark data poses[7]

- **Legal and regulatory risk.** If data covered by mandate or regulation – such as confidential, financial information (credit card or other account data) or sensitive

[6] https://www.gartner.com/it-glossary/dark-data

[7] https://www.cio.com/article/2686755/data-analytics/the-dangers-of-dark-data-and-how-to-minimize-your-exposure.html

(personal) data – appears anywhere in dark data collections, its exposure could involve legal and financial liability.

- **Intelligence risk**. If dark data encompasses proprietary or sensitive information reflective of business operations, practices, competitive advantages, important partnerships and joint ventures, and so forth, inadvertent disclosure could adversely affect the bottom line or compromise important business activities and relationships.
- **Reputation risk**. Any kind of data breach reflects badly on the organizations affected thereby. This applies as much to dark data (especially in light of other risks) as to other kinds of breaches.
- **Opportunity costs**. Given that the organization has decided not to invest in analysis and mining of dark data by definition, concerted efforts by third parties to exploit its value represent potential losses of intelligence and value based upon its contents.
- **Open-ended exposure**. By definition, dark data contains information that's either too difficult or costly to extract to be mined, or that contains unknown (and therefore unevaluated) sources of intelligence and exposure to loss or harm. Dark data's secrets may be very dark and damaging indeed, but one has no way of knowing for sure. This can't cultivate complacency or indifference in those who contemplate those risks at all seriously.

As the general trend is for firms increasingly collect more information, they need to consider the costs of data which are dangerous and unprofitable to what they are doing.

3.3.2 Data acquisition cost

Acquiring data to be used or to be "sold" has a cost. Companies that seriously want to integrate the use of big data in their operations (or to develop a major product) must be willing to think about the costs of acquisition and maintaining the "supply". (Related to this topic is also the discussion on the cost of personal data, which is dealt with further in the Chapter 7.)

Companies will need to protect themselves from changes in regulations, their competitors, agreements with their suppliers, and potentially even the changes that affect the supply chain of their suppliers. Recent regulations on privacy and protection of personal data, while beneficial for owners of personal data, may upend the business models of certain companies or at least increase the costs of the business. This is not unusual for the private sector. But, it may hurt the investment outlook, if the industry is insecure about its continued growth or the regulatory landscape.

Whether the data is provided for "free", the company would generally need to offer something in exchange for the data, except for most open data cases. This may come in the form of providing services in exchange, providing data in exchange, partaking in a platform, or adopting certain technical standards. Hence, the company will need to evaluate carefully, whether the benefit will exceed the costs of using big data. This will need to be evaluated within the business model itself.

3.3.3 Systems acquisition and operation

While BDA can prove to be profitable to a firm in the transport sector, the firm must consider the costs and resources that a successful BDA usage requires. The firm (or a partner, or the government) must invest technical systems that can handle big data. This starts from the collection to communication to storage and retrieval systems for big data. The cost of investment and operation can be high. For instance, an estimate puts a petabyte Hadoop cluster to cost about $1 million annually[8]. To put this in perspective, it is estimated that about 0.5 petabytes of data were collected by every automotive manufacturer in 2013, and that it is expected to increase to 11.1 petabytes annually by 2020 (Khan, Rahman, et al. 2017). This means that hypothetically that any firm in 2013 that wanted apply store big data from automotive manufacturers would have to be paying $0.5 million annually, with increasing costs. A data warehouse could cost between $1 to $25 million per petabyte annually[9].

Also note that Hadoop is a collection of open-source software; enterprise software providers would probably cost even more, although that may be accompanied with better services and performance (Nik Rouda 2014). Indeed a major question is whether the system should be built by the firm or provided perhaps as a service, leading to some business models such as Hadoop-as-a-Service.

3.3.4 Re-organization and adaptation of processes

As innovations are adopted and business models are modified, “organizational restructuring often occurs” (Bucherer, Eisert, and Gassmann 2012) and in fact may even be a prerequisite for a successful business innovation process. There are “costs, time and difficulties” that should not be underestimated.

One example would simply be in the moving towards digitalization of processes, with a clear line to ensure the **quality** of data collected along the process. For this, it would be important to question whether there are sufficient **quality assurance mechanisms** implemented along the data value chain. Fu and Easton (2017) suggests that decision makers redesign processes to ensure the data meets a clearly defined level of quality from the first collection and subsequently in transmission, storage and analysis; develop and support employees in treatment of the data with appropriate tools, and to monitor the changing value of data. This can support the avoidance of increasing dark data and bad data. Suitable industry standards may need to be developed, to supplement existing data quality standards[10], and evaluation methods that can deal with big data (De Groeve et al. 2014; Hubauer et al. 2013).

Another aspect is the **cost of compliance with legal requirements**, such as the General Data Protection Regulation (GDPR), or other policies that require auditing data. The cost of *non-compliance* will certainly exceed the cost of compliance (see Figure 5 for a summary of

---

[8] https://www.forbes.com/sites/ciocentral/2012/04/16/the-big-cost-of-big-data

[9] https://aws.amazon.com/blogs/aws/amazon-redshift-the-new-aws-data-warehouse/

[10] ISO has produced standards for ensuring industrial data quality from a process point of view. (https://www.iso.org/committee/54158/x/catalogue/p/0/u/1/w/0/d/0)

penalties), however that does not negate the high economic cost that any firm dealing with data, particularly sensitive and personal data, will have to bear. The main costs[11] will be in the "auditing and classifying the data". These are processes that cannot be excluded in the compliance of GDPR, since it will identify the risks, facilitate information sharing, support consent evaluation, enable right of request for copies of data or to have data deleted, etc. Furthermore, it will require in some cases hiring new staff, such as a Data Protection Officer.

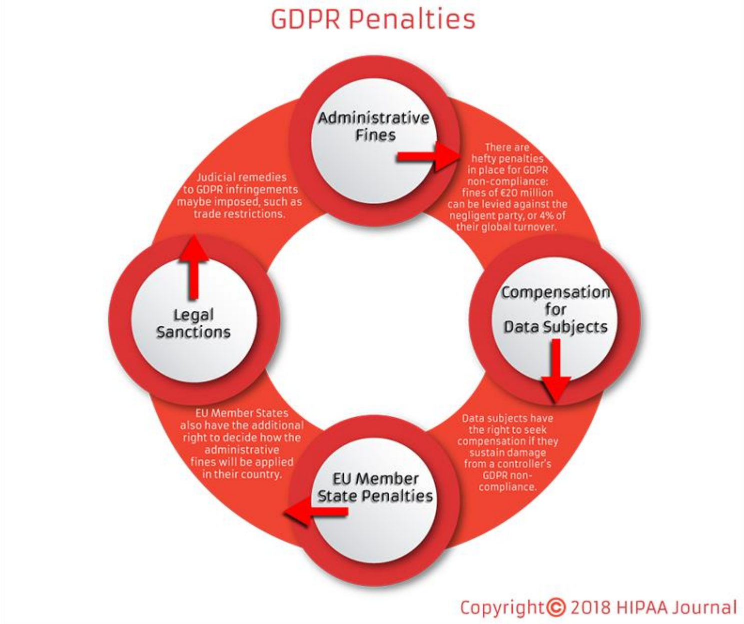

Figure 5 Penalties imposed by the GDPR

In addition to staff for compliance and quality assurance, the firm will basically also need to pay for staff (either in house or outsourced) to run BDA and create value out of it. This may also require a period to experiment, research and refine methods. Therefore in short there is considerable investment in the process of creating value that the firm must account for. This includes the issue of finding experienced staff to spearhead big data innovation projects in the yet very traditional industry of transport. In this sense, start-up companies have the advantage of in a sense skipping the costly and challenging adaptation process.

3.4 Summary

In this chapter we have examined some of the influences big data, and in particular the analytics of it, will have on the business model and operations of the participating firm. In particular, BDA can help transport operations improve in terms of:

- increasing the SSA of the transport sector actors;
- improving the capacity of transport network, either through planning, real-time management, and mitigating current risks of disruption;

---

[11] https://www.hipaajournal.com/the-cost-of-gdpr-compliance/

- improving the provision of transport services, through better marketing activities, vehicle routing, and better service provision and scheduling; and
- supporting the shift towards sustainability.

Transport sector actors are also expected to take a part in the role in the data economy as a data supplier. While data collected in the transport sector will primarily support the activities of other transport sector actors, it is worth exploring how this data can be used in other segments of the economy, such as in advertising, research, and the public sector.

While we expect many benefits, it ultimately depends on whether the firm can pull it off successfully. The firm will also have to consider various costs that accompany their foray into using and being a supplier in the data economy. Identifying the risks and the costs that apply to each individual company will increase their chances of making the digital economy strategy of the EU successful.

4 Industry Level: Growth of an Industry using Big Data Technology

In this chapter the data users, suppliers and facilitators are discussed on an industry level by describing EU's big data technological innovation system. A technological innovation system is a concept used to analyse the innovation ecosystem of a particular market (M. Hekkert et al. 2011). In an innovation system, the interaction between different actors in relation to contextual factors (such as institutions, technologies and policies) determines the creation of successful products and services. According to (M. Hekkert et al. 2011), "The concept of the innovation system stresses that the flow of technology and information among people, enterprises and institutions is key to an innovative process. It stresses the interaction between actors who are needed in order to turn an idea into a successful process, product or service in the marketplace". Mapping the structure of a technological innovation system is done by describing the actors, technological factors, and policies[12].

The goal of this chapter is to describe the EU big data innovations market to assess the potential for big data innovations in transport. The issues of the presence of dominant players in the data market is discussed separately in chapter 8. Lastly, externalities of the data market are described.

4.1 Actors: Growth of data suppliers, data users and the data economy

In the framework of EU's policy initiative 'towards a digital market', effort was done to map the EU's data market. This has led to the creation of the European Data Market Monitoring Tool[13]. The monitoring tool and accompanying study reports (IDC and Open Evidence 2017; Cattaneo et al. 2018) give an overview of the size of EU's data market (i.e. the exchange market of products and services derived from data) by making an inventory of data company suppliers, data users and data workers. In addition, an estimation of the data economy (i.e. the economic impact resulting from the data market) is given, as well as a comparison of EU's data market with the data market of other countries, most notably the US. The description of the structure of data economy given below is based on those reports.

4.1.1 Data companies: suppliers and users

In the data market, there are more data users than data suppliers (Table 1). However, the number of data suppliers is increasing faster. The penetration rate of data user companies varies across the EU. Countries with more data orientated industries, such as professional- and financial services and retail, tend to have a higher data company penetration rate. In 2017 the revenues of data companies made up 3.2% of the total revenues of companies based in Europe. Data revenues are expected to keep rising. Forecasts show a yearly revenue increase of 8.3% up to 2025. Although smaller Member States show a high long-term growth, companies in the

[12] The fourth factor making up a technological innovation system are 'Networks'. Hekkert et al. (2011) note that networks are difficult to grasp and are often implicit. As only significant effort reveals big data transport networks, we decided to omit the fourth component. The other three, actors, technology and institutions, suffice to obtain a general picture.

[13] Available at http://datalandscape.eu/european-data-market-monitoring-tool

larger Member States will have the biggest contribution to the Data Economy in the upcoming years.

Table 1 Data supplier companies and data user companies in EU for 2016 and 2017

	2016	2017	Growth
Data suppliers	261,000	276,000	5.7%
Data users	676,000	690,650	2.1%

The potential growth of data users in an optimistic scenario could amount up to 853,000 companies by 2025. These companies are using data for production and/or business processes. By monetising the use of data within their operations, companies create added value. This trend is recognised by an increasing number of organisations. By investing in tools and technology that allow (improved) mining/collecting, processing and analysing data organisations are adopting the concept of "data as a service" as a business model. This puts them at the frontline of the digital transition as they apply data-driven decisions and explore new areas of data application.

4.1.2 Data workers

Table 2 shows the statistics of data professionals in the EU. In 2017 the number of data professionals grew significantly. The share of data professionals within the total labour market is expected to grow, indicating the increasing demand for data workers. Under optimistic conditions, the number of data professionals could grow at a yearly rate of more than 10% in the upcoming period. Innovative markets, such as the data market, generate new jobs.

Table 2 Data workers in the EU for 2017

	Data professionals	Increase data professionals	Share of the total workforce	Share of the total workforce (2020)	Excess demand data professionals
EU28	6.7 million	8%	3.2%	3.6%	449,000

Data professionals usually work in high-skilled, multi-disciplinary teams. Despite the creation of new jobs, there is currently an unbalance in the demand and supply for data related positions. In 2017 there were 449,000 data related unfilled positions in the EU28. This data skills gap has also been present in the previous years. This structural problems is expected to continue. This gap could grow to 1 to 2 million unfilled positions by 2025. Actions relating to reducing the mismatch between the skills required and the skills offered (lack of data-scientific educated personnel) and improving market efficiency could help reduce the data skills gap.

4.1.3 The Data Market and the Data Economy

The **data market** is the place where data-driven products and services are exchanged. Benefits from data driven innovation require complex digital processes and intelligent thinking. There are several countries within the EU that have been home to these successfully operating companies for a number of years. Combined with a lower share of traditional small- and medium enterprises, countries like the U.K., Netherlands and Germany experience relatively increased benefits from the data market.

In 2017 the data market was valued at 65 billion euros, growing at a rate of 9.3% compared to 2016. With an increasing demand for data-as-a-service, the data market is estimated to grow at an average rate of 6.0% to 2020. Most consumers of data technologies are found in the bigger industries. Forecasts indicate that manufacturing, finance, professional services, IT and retail industries are responsible for the main share of the data market growth in 2025.

The **data economy** is defined by the data market and its direct/indirect (and induced) effects on the economy as a whole. This means the development of the data market directly correlates with the development of the data economy. Both are projected to show growth in all Member States. Impacts on other industry can specifically be found in the manufacturing, finance, professional services, IT and retail industries since these industries increasingly rely on data as a service.

A steady pace of innovation and a good economic climate ensured an increased growth of the European data economy, and the data economy of each of the individual Member States. This resulted in a sectoral growth that was almost 5 times larger than the general economic growth of the EU28. The value of the data economy is estimated at 335.5 billion euros (2017), growing with 11.8% compared to last year. In 2017 the data economy makes up 2.4% of EU's GDP. Under an optimistic scenario, estimates haven shown that the overall impacts of the data economy could be more than 1 trillion euros by 2025.

4.1.4 Other data markets: USA, Japan and Brazil

Considering the data markets of the USA, Japan, Brazil and Europe, the USA data market is the largest data market in the world, and with that, the most dynamic one. The European data market is catching up with the market leader, especially in the area of data suppliers which grew more than 2 times the speed of the USA in 2017. Regarding Europe's data economy, direct and indirect effects increased at a higher speed than in other markets as well. This resulted in an increase in the Data Economy as a share of GDP. A disadvantage for non-USA markets is related to structural factors (such as the higher presence of small- and medium enterprises), and cultural and educational factors (ability to create and keep data-related skills). Although Europe may lack the presence of giant IT companies opposed to the USA, there is a healthy presence of digital startups and innovation capacity. Other digital technologies could help create opportunities within the European market.

4.2 Technology: growth of data facilitator companies and accompanying issues

The goal of this section is to understand the growth of big data technology driven companies and communities in the EU. We analyse the big data open-sources software communities and vendors with focus on their origin. The communities and vendors are not necessarily serving the transport sector.

The landscape of big data technologies is highly complex as shown in Chapter 3 of LeMO Deliverable 1.3. For instance, the quantity of technologies are surging; some are open-sources but not all; there is no consensus among studies for a universal taxonomy of the big data technologies.

Open-source software is a type of computer software of which the source code is released under a license in which the copyright holder grants users the rights to study, change, and distribute the software to anyone and for any purpose (Laurent, 2004). According to Jay Kreps[14], CEO of Confluent, a company contributing the popular big data open-source tool Apache Kafka messaging system, we are at a unique juncture in the trajectory of open source relative to the tech ecosystem as a whole. It is not exaggerated to say that nearly all Big Data technologies are open-source.

Big data platform vendors design their own big data stack based on big data open-source software. Vendors are providing not only software to the market, but support and services as well. Table 3 shows six major big data vendors. The first three vendors provide a downloadable software and others offer a cloud service. Some of well-known vendors, such as Amazon Web Services, Google Cloud Platform, IBM BigInsights, and Microsoft HDInsight, were excluded because their products are not bound to use big data stack.

Table 3 Big data Platform Vendors

Vendors	Headquarters	Type	Number of employees	Revenue (USD)	Founded
Cloudera[15]	Palo Alto, California, US	Public	1,600+	390.6M[16]	2008
Hortonworks[17]	Santa Clara, California, US	Public	500+	284.9M[18]	2011
MapR[19]	Santa Clara, California, US	Private	500+	118.8M[20]	2009
Qubole[21]	Santa Clara, CA, US	Private	200+	18.8M[22]	2011
Altiscale[23]	Palo Alto, California, US	Private	50+	2.4M[24]	2012

Cloudera is a frontier company which offers a commercial big data platform and is the most popular vendor to its competitors. Cloudera was founded by three engineers from Google, Yahoo! and Facebook and a former Oracle executive. Of the vendors, Horton offers only open-

14 https://www.datanami.com/2018/02/26/weighing-open-sources-worth-future-big-data/
15 https://www.linkedin.com/company/cloudera/
16 https://www.cnbc.com/quotes/?symbol=CLDR
17 https://www.linkedin.com/company/hortonworks/
18 https://www.cnbc.com/quotes/?symbol=HDP
19 https://www.linkedin.com/company/mapr-technologies/
20 https://www.owler.com/company/mapr
21 https://www.linkedin.com/company/qubole/
22 https://www.owler.com/company/qubole
23 https://www.linkedin.com/company/altiscale/
24 https://www.owler.com/company/altiscale

source Hadoop technology without modifications. The initial chief executive from Hadoop previously worked at Yahoo!. Unlike Cloudera and Hortonworks, MapR is privately held company and relatively smaller than competitors in terms of the size. MapR executives originate from six different companies: Google, Lightspeed Venture Partners, Informatica, EMC Corporation and Veoh. Qubole is a big data-as-a-Service (BDaaS) provider founded by two Facebook engineers. Altiscale provides another solution of BDaaS founded by an engineer from Yahoo!.

It is obvious that *the US is leading the game in big data technologies.* All vendors are located in the US, and the previous affiliation of all founders is also located in the US. In Table 3 it is noteworthy to pay attention on the last vendor, Altiscale. Altiscale was founded in 2012 in the US and then it was acquired by SAP in September 2016. This acquisition opens the door for SAP into the big data World.

SAP is a multinational German software corporation. According to Forbes, SAP is the largest company in Software and Programming section in Europe and the third largest company in the world[25]. Unlike other largest Internet companies like Google, Amazon or IBM, SAP enhances its cloud portfolio to market by acquiring a big data solution company. This raises an important question: Why does the third largest software company in the world needs to acquire a big data solution instead of developing its own solution?

One of the possible answers is the lack of skilled professionals with big data technical knowledge. It is in fact also introduced as a core challenges of big data in Transportation in LeMO Deliverable 1.1 (Hee, et al., 2018). Figure 6 illustrates the number of big data degree program across the country of origin in 2016. It is obvious that the quantity of big data degree programs in the EU falls short of that in the US. More than half of study programs (55%) are available in the US, followed by UK (16%). Even when aggregated, the number of degree programs in the EU does not reach US levels.

Fortunately, many politicians in the EU are aware that big data technologies in the EU lags behind the US. According to André Loesekrug-Pietri[26], a German-French investor, "We are very worried about how slowly things are moving forward". Stefan Heumann[1], the co-director of Berlin-based think tank Stiftung Neue Verantwortung (SNV), cautioned "a government has to be judged by its actions ... it also has to include implementing a range of policy initiatives, from developing data pools to training programs for workers."

One of the primary functions of the EU in this area is to support research. They aim to foster research that can serve in driving economic growth and creating jobs, ensuring Europe's future global competitiveness. One of the illustrative project in this field is the EDISON project[27], short for *E*ducation for *D*ata *I*ntensive *S*cience to *O*pen *N*ew science frontiers. It is aimed at defining a data science body of knowledge, establishing a model curriculum and ultimately promoting

[25] https://www.forbes.com/global2000/list/#industry:Software%20%26%20Programming, Retrieved: 31.07.2018

[26] Source: https://www.politico.eu/article/germany-falling-behind-china-on-tech-innovation-artificial-intelligence-angela-merkel-knows-it/

[27] http://edison-project.eu/

big data education in EU. *Figure 6* shows the number of big data degree programs according to their survey[28] (EDISON, 2016). It shows that from an education standpoint, the USA stands out the most, with over 45% of the degree programs surveyed.

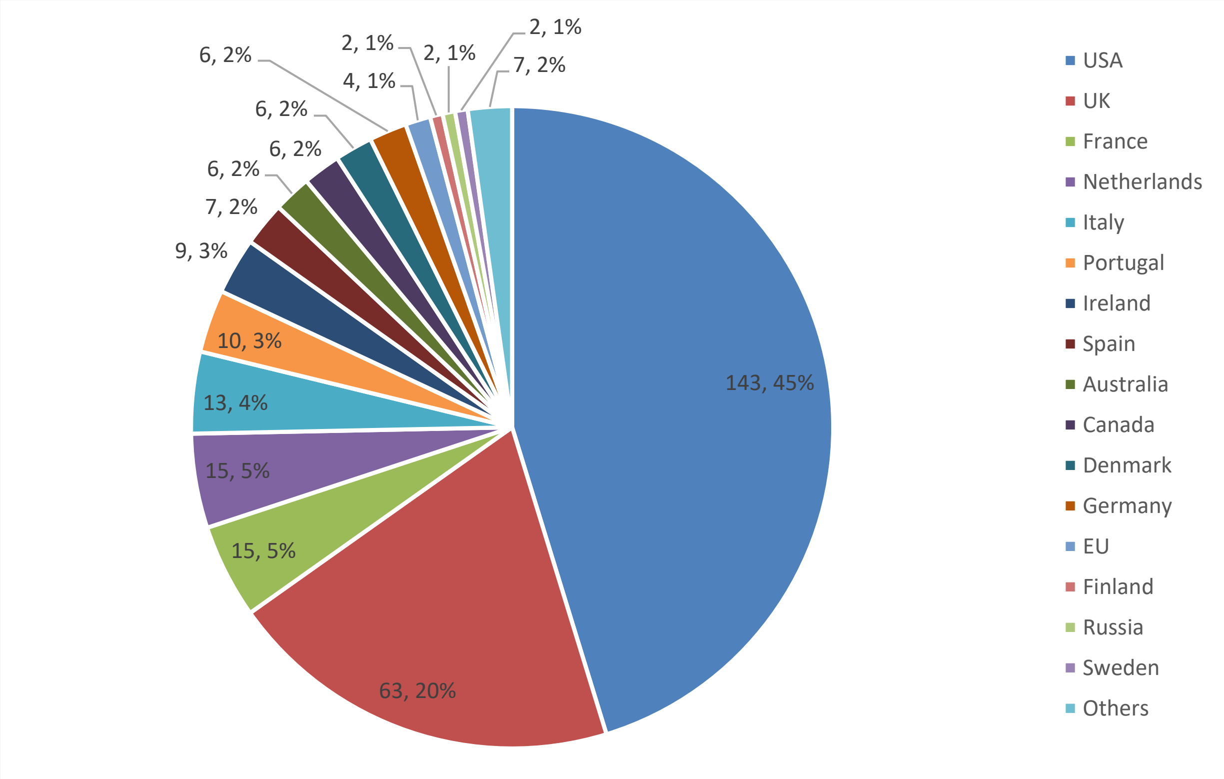

Figure 6 Country of origin for big data degree programs (EDISON, 2016)

Another reason the US is leading the big data technology is possibly attributed to the investment spending on Information and Communication Technology (ICT). The ICT sector contributes to technological progress, enhances the output and productivity, and impacts on employment and other parts of the economy. The ICT investment among the OECD countries from 2001 to 2010 is visualized in Figure 7. It is measured as a percentage of total non-residential gross fixed capital formation. It indicates the policy in the selected countries and which industry each country wants to foster. According to Figure 7, the US shows the strongest ICT investment among the OECD countries. The US is leading the stage in both on big data education and investment on ICT. Europe needs to take an immediate action to catch up the US, otherwise, Europe will not take a leading role in the big data technology domain.

[28] The survey was conducted using an internet search and questioning the project members. The statistical figures should not be treated as conclusive, but indicative. The countries surveyed were

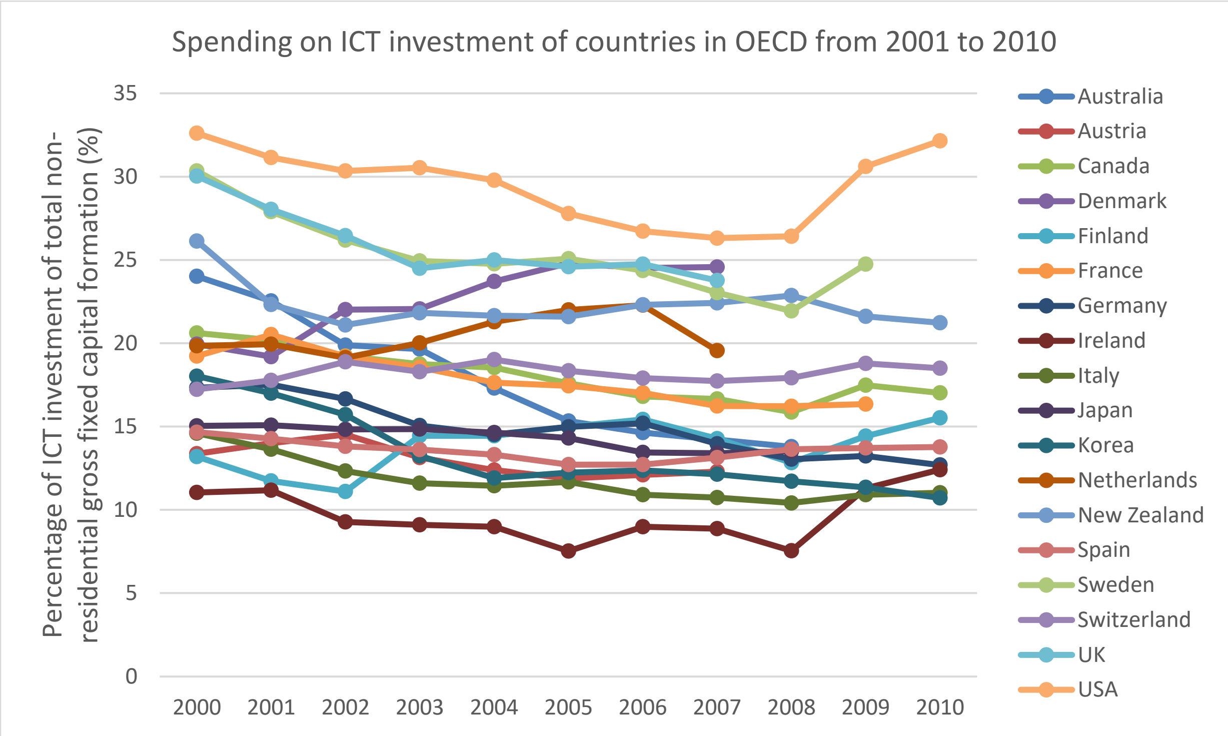

Figure 7 Spending on ICT investment in 18 countries in OECD, Percentage, 2001-2010[29]

4.3 Policies: role of public agencies in regulating big data in the transport sector

As discussed previously, the transport sector is highly regulated, because public agencies perceive a responsibility to mitigate transport externalities and to ensure the strategic role of the transport sector for the economic and general well-being of their constituents.

Widely discussed and partly covered in Deliverable 1.2 are the policies by the EU and other countries on what is expected from the application of big data in the transport sector. The main thrusts of EU policy relevant in this sphere are summarized below.

The EU stake in the research and policy big data are two-fold: unlocking the commercial benefits of a big data, and anticipating negative impacts and regulating to protect EU common values.

The commercial benefits of big data are summed up in the following directions:

- Supporting the development of technologies and the start-ups/firms that can support the Digital Single Market[30],
- Developing a Data Economy[31] which complements the Digital Single Market, and

[29] OECD (2018), ICT investment (indicator). doi: 10.1787/b23ec1da-en (Accessed on 17 August 2018)

[30] https://ec.europa.eu/commission/priorities/digital-single-market_en

[31] https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy

- Developing and improving the transport industry (whether owned or operated by the private or public sector).

The **Digital Single Market** aims to reduce barriers within the EU, between Member States for online services to develop. The main strategies are presented in Box 1. The development of a **Data Economy,** on the other hand, aims to "unlock the re-use potential of different types of data and its free flow across borders". Whether using big data analytics or traditional analysis methods, the aims of the Data Economy are to:

- transform Europe's service industries by generating a wide range of innovative information products and services;
- increase the productivity of all sectors of the economy through improved business intelligence;
- better address many of the challenges that face our societies;
- improve research and speed up innovation;
- achieve cost reductions through more personalized services;
- Increase efficiency in the public sector.

Box 1: Aims of the Digital Single Market[32]

Legal

- Boosting e-commerce in the EU by tackling geoblocking, making cross-border parcel delivery more affordable and efficient
- Modernising the EU copyright rules to fit the digital age
- Updating EU audio-visual rules and working with platforms to create a fairer environment for everyone, promote European films, protect children and tackle hate speech
- Unlock the potential of a European data economy with a framework for the free flow of non-personal data in the EU
- Adapting ePrivacy rules to the new digital environment

Technical/technological

- Scaling up Europe's response to cyber-attacks by strengthening ENISA, the EU cybersecurity agency, and creating an effective EU cyber deterrence and criminal law response to better protect Europe's citizens, businesses and public institutions
- Ensuring everyone in the EU has the best possible internet connection, so they can fully engage in the digital economy, the so-called "connectivity for a European gigabit society"
- Helping large and small companies, researchers, citizens and public authorities to make the most of new technologies by ensuring that everyone has the necessary digital skills, and by funding EU research in health and high performance computing

[32] https://ec.europa.eu/commission/priorities/digital-single-market_en

It has been suggested that big data could also unlock further potential of traditional sectors, such as the transport sector. For the transport industry, following the principles of sustainable transport is an objective of increasing importance[33]. Data could serve as a means to support this. Simultaneously, the goals of the Data Economy could be realized.

Key developments needed according to the Transport White Paper are summarized in Box 2. Big data will have only an indirect influence on Goals 1 to 6. However, for Goals 7 to 10, big data, the digital single market, the data economy is expected to have a significant influence. For instance, big data techniques can be expected to integral in the reaching the full potential benefits of using the various traffic management systems, such as the Single European Sky Air Traffic Management Research (SESAR), European Railway Traffic Management System (ERTMS), Intelligent Transport Systems (ITS), the SafeSeaNet (SSN), Long-range identification and tracking (LRIT) and River Information Services (RIS).

Box 2: Ten goals for a competitive and resource efficient transport system: benchmarks for achieving the 60% GHG emission reduction target[34]

Developing and deploying new and sustainable fuels and propulsion systems

1) Halve the use of 'conventionally-fueled' cars in urban transport by 2030; phase them out in cities by 2050; achieve essentially CO2-free city logistics in major urban centres by 2030.
2) Low-carbon sustainable fuels in aviation to reach 40% by 2050; also by 2050 reduce EU CO2 emissions from maritime bunker fuels by 40% (if feasible 50%).

Optimizing the performance of multimodal logistic chains, including by making greater use of more energy-efficient modes

3) 30% of road freight over 300 km should shift to other modes such as rail or waterborne transport by 2030, and more than 50% by 2050, facilitated by efficient and green freight corridors. To meet this goal will also require appropriate infrastructure to be developed.
4) By 2050, complete a European high-speed rail network. Triple the length of the existing high-speed rail network by 2030 and maintain a dense railway network in all Member States. By 2050 the majority of medium-distance passenger transport should go by rail.
5) A fully functional and EU-wide multimodal TEN-T 'core network' by 2030, with a high quality and capacity network by 2050 and a corresponding set of information services.
6) By 2050, connect all core network airports to the rail network, preferably high-speed; ensure that all core seaports are sufficiently connected to the rail freight and, where possible, inland waterway system.

Increasing the efficiency of transport and of infrastructure use with information systems and market-based incentives

7) Deployment of the modernized air traffic management infrastructure (SESAR) in Europe by 2020 and completion of the European Common Aviation Area. Deployment of equivalent land and waterborne transport management systems (ERTMS, ITS, SSN and LRIT, RIS). Deployment of the European Global Navigation Satellite System (Galileo).

[33] https://sustainabledevelopment.un.org/topics/sustainabletransport
[34] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52011DC014

8) By 2020, establish the framework for a European multimodal transport information, management and payment system.
9) By 2050, move close to zero fatalities in road transport. In line with this goal, the EU aims at halving road casualties by 2020. Make sure that the EU is a world leader in safety and security of transport in all modes of transport.
10) Move towards full application of "user pays" and "polluter pays" principles and private sector engagement to eliminate distortions, including harmful subsidies, generate revenues and ensure financing for future transport investments.

4.4 Externalities of big data technologies

The usage of big data technologies can affect individuals or groups that were not involved with the production or consumption of the technology. In economics, the impacts on third parties are referred to as externalities. Most externalities are negative, such as air pollution coming from a power plant or visual pollution caused by the construction of new buildings. Other externalities are positive, such as companies benefiting from the high education of their workforce.

Various big data externalities relate to social or ethical issues, such as loss of autonomy or privacy infringement. These issues are explained in more detail in D2.3 and 2.4 and in the chapter dedicated to privacy concerns, chapter 7. An example of such a negative externality is that big data is used to infer about a population, leading to assumptions being made about individuals who were not part of the group of whom data is collected.

Another major externality is the effect that the big data economy has on the environment in terms of the energy consumption that is required for collecting, storing and processing data (Milne and Watling 2018). In 2013, electricity required for ICT infrastructure uses 10% of global energy consumption, and as ICT technologies keep developing and spreading, the share is expected to grow rapidly (Mills 2013). A data centre uses 100 to 200 times more electricity per square meter than a modern office building (Mills 2013). *Figure 8* shows the estimated energy consumption of 'the Cloud' (i.e. the entire ICT ecosystem) compared to the illumination industry and the Electric Vehicle industry (EV). The amount of coal required represents 40% of the total energy that is required and is based on the current relative share of coal.

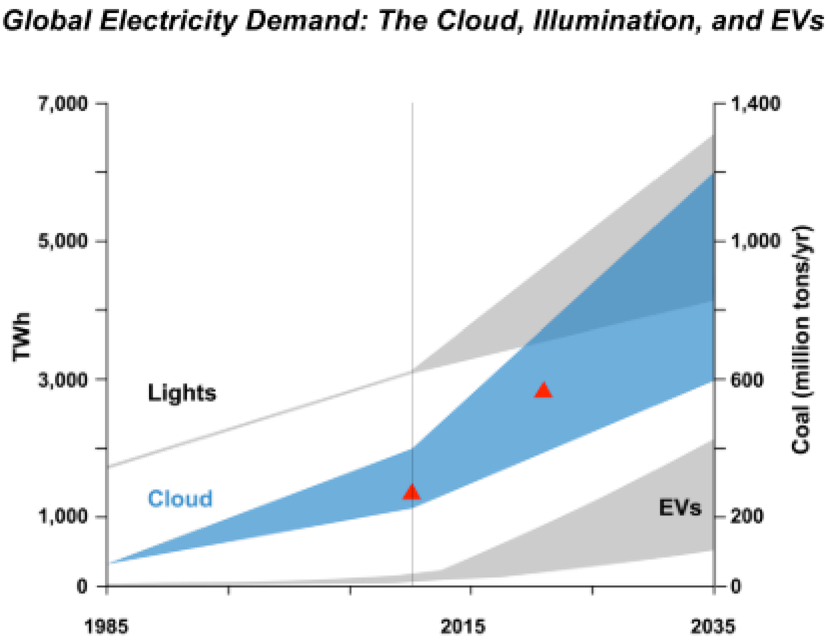

Figure 8 Comparing the global electricity demand of the ICT sector with the illumination sector and electric vehicles (EV) sector. Source: (Mills 2013)

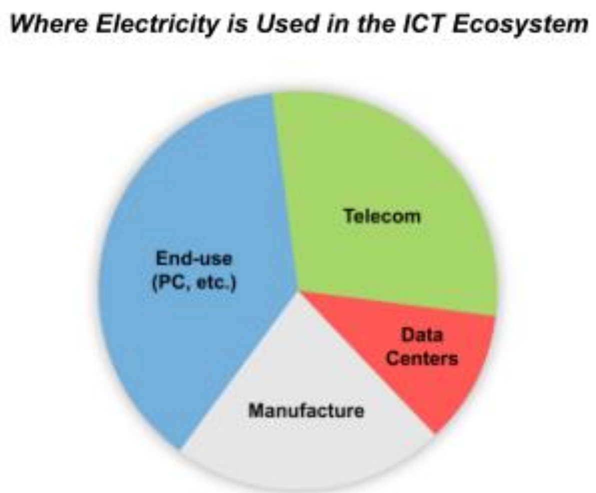

Figure 9 Four main components of the ICT ecosystem and relative share in power usage. Source (Mills 2013)

The ICT ecosystem has four main components and their relative share in electricity consumption is shown in *Figure 9*. The four components are:

- Data centres responsible for storing, transferring and processing information
- Manufacturers that produce all ICT equipment
- Telecom refers to the wired and wireless networks used for transferring data, such as WiFi, cellular and fibre

- End user devices such as mobile phones, PC's and tablets used at home, in offices and in factories

Maintaining its current growth, it is likely that the big data infrastructure will put an equal amount of pressure on the environment as transportation is currently doing (Milne and Watling 2018), requiring the need for energy efficient innovations. One way for more efficient usage of energy is to use the heat surplus generated by data centres for heating residential houses, greenhouses and office buildings (UNICA 2017).

4.5 Summary

The factors influencing the growth and innovation in Europe's data economy are considered in this chapter. The EU's data economy is described by using the notion of a technical innovation system (Hekkert et al. 2011). In a technological innovation system, the interaction between actors, technologies and institutions determines the creation of successful products and services. Each of the components of EU's big data innovation system are summarized here. In addition, economic externalities are discussed.

Actors:

- The amount of data companies and data users in the EU are growing quickly, while a shortage in data workers remains. The size of the data market is expected to grow annually by 6% until 2020. Optimistic estimates suggest the size of the data economy to be €739 billion in 2020 or 4% of EU GDP, from €300 billion or 2% of EU GDP in 2016.
- The US data market outperforms the EU market in terms of size and dynamic, but the EU is catching up, in particular for the number of data suppliers, which grew with twice the speed of the US market in 2017.
- The EU will continue to experience a lag due to the US regarding structural factors (fewer data SME), cultural/educational factors (ability to create and keep data-related skills), and the presence of IT giants. However, there is a healthy presence of digital start-ups and innovation capacity.

Technologies:

- Nearly all (5 out of 6) big data vendors are located in the US, however since the technologies are open source, they remain accessible to the EU market.

Policies:

- The EU is doing significant effort to advance EU's big data economy through implementing policies developed under 'Digital Single Market' framework.
- Specific targets are set for big data applications in various traffic management systems, such as the Single European Sky Air Traffic Management Research (SESAR), European Railway Traffic Management System (ERTMS), Intelligent Transport Systems (ITS), the SafeSeaNet (SSN), Long-range identification and tracking (LRIT) and River Information Services (RIS).

Actors, technologies and institutions reflect the main structure of EU's big data industry. However, for a full assessment of EU's big data ecosystem, research on the *performance* of the structure is required. This can be conducted according to the 7 functions of a technological innovation system (M. P. Hekkert et al. 2007):

1. entrepreneurial experimentation
2. knowledge development
3. knowledge exchange
4. guidance of the search
5. formation of markets,
6. mobilization of resources
7. counteracting resistance to change

Externalities:

- Big data usage in the economy causes negative externalities regarding the pressure on climate change. Currently, the IT ecosystem takes up more than 10% of the world's energy consumption. As the IT sector keeps growing, in particular with the advent of electric vehicles, we need to consider more efficient energy usage of data centres, IT manufacturing, end-use devices and IT infrastructure.

5 National Level: Application of Big Data in the Governance Process

The aim of this chapter is to reveal a part of big data that is relevant to government endeavours. The first part elaborates on general societal issues, the second part on transport specific issues. This chapter focusses on the following roles of the government: internal organisational efficiency, service provision and policy making. Another role of the government is making legislation. This topic is not considered here, as it was already treated in D1.2 on 'Big Data Policies'. Moreover, one of the most pressing issues on using big data relates to privacy and personal data protection. Due to its significance and the interconnectedness to issues relevant to the private sector, this topic is considered in a separate chapter. In addition, this chapter only focusses on domestic politics. Foreign politics constitutes a different dimension with different issues and is likewise treated in a separate chapter.

5.1 Big data governance by governments: general issues

Governments use big data to improve their operations in a similar way to how businesses apply big data to improve theirs. But the way in which big data is utilized reflects different goals and values. There are fundamental differences between businesses and governments in goals and organisational structure (Kim, Trimi, & Chung, 2014). Businesses typically use BDA to minimize costs and maximize shareholder value to gain a competitive edge over competitors. Competition is the driving force for big data-based innovations in a fast-paced environment with hierarchical structures. For governments, the primary goal of BDA is more ambiguous and contested: adding public value. Generating public value goes beyond simply improving efficiency and effectiveness, and includes values such as democracy, public participation, legitimacy and inclusion (Bryson, Crosby, and Bloomberg 2014). Many actors are involved in long and thoroughly checked decision making processes. Governments use BDA in many ways to facilitate the act of governing, such as in job creation, welfare distribution, crime prevention, fraud prevention, improving education or the healthcare system, in national security and so on.

In this section, the application of big data by governments is explained in two ways. First, the organisational structure of governments is considered by answering the question how governments can use BDA in its own institution. The second part deals with how governments apply BDA in governing, specifically in public service provision and policy making.

5.1.1 Internal Governance: big data users

The successful application of big data by governments depends on the extent to which the opportunities of big data are exploited and the challenges are dealt with. This is determined by the 'data-culture' within an organisation, which refers to the capacity of individual employees and the organisation to exploit big data in such a way that it is beneficial for the organisation (Giest, 2017, p. 389).

Various strategies exists that give guidance to incorporating big data in the governance process. One approach is described in detail here, the Big Data Governance Blueprint (Kim & Cho, 2017). The main argument is that from an organisational point of view, exploiting big data requires a strategic approach, in which objectives are clearly defined and the challenges of big data, such as data quality, privacy and security, are dealt with within an appropriate organisational structure. The big data governance strategies are applicable to any kind of institution, be it in the private or public sector. However, as aims and organisational structure differs, various components in the governance process take up various shapes.

Kim and Cho (2017) propose a big data framework (see *Figure 10*) in which four aspects need to be considered related to objective, strategy, components and IT infrastructure.

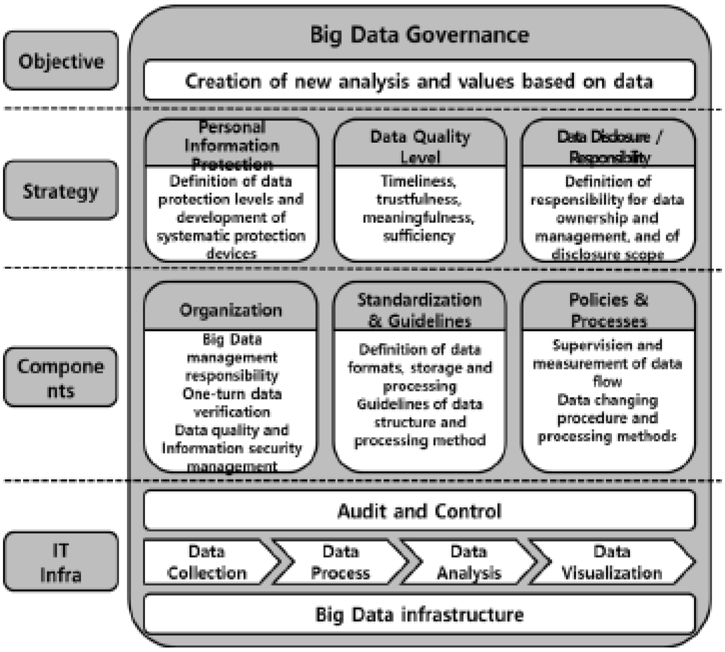

Figure 10 Big data governance framework for governments (Kim & Cho, 2017)

Objectives: in the planning phase, a clear definition of the aim is required to ensure that big data is used effectively. The outcome of big data analysis is only as good as carefulness with which the data is processed. A bad data collection approach and bad data analysis leads to bad outcomes. Crucial for success is firstly, an understanding is required of what questions big data can answer and what big data cannot, and secondly, how the big data will answer the question. Big data in itself isn't useful – it is only as useful as the skill of those that are handing the data and design the data process.

Strategy: during the data process, issues of privacy, data quality and data disclosure and responsibility need to be managed.

- Privacy is a sensitive issue. Not every kind of data can be collected, and collected personal information requires sufficient protection. A strategy needs to be in place to prevent data breaches and access to data of unauthorized individuals.
- Data quality attributes may be classified to timeliness, trustfulness, sufficiency and meaningfulness. **Timeliness** is necessary as data can age quickly and the continuity of the whole process needs to be maintained. **Trustfulness** refers to the reliability of the data source. **Meaningfulness** concerns whether the data collected helps with answering the question that is posed in the beginning. **Data sufficiency** is achieved when it is good

enough to derive adequate conclusions. All four attributes help with collecting right data that is required to lead to the right conclusions.

- The last strategical component ensures that it is clear how and by who the data process is managed and disclosed.

Components: factors that enable the right strategy are the organisation, standardization and guidelines, policies and processes. Having an overview of the components enables making the right strategy to achieve the objectives. These are

- The organisational structure refers to all the different components that are part of reaching the objective, such as the management, policy makers, the IT team, and the data security team. It refers to the roles and skill of each team member.
- Standards and guidelines regarding storage and processing need to be defined to maintain a smooth communication and data transfer between the different stages of processing the data.
- Policies and procedures operate between and within different parts of the organisation and function along various standards and guidelines. Defining certain procedures, such as supervision, data handling, data sharing, management, and team meetings contributes to an efficient functioning of the whole organisation

IT infrastructure: the technical part of big data is determined by the IT infrastructure, such as the available hardware and software for each data stage: data collection, processing, analysis and visualization. Using big data requires resources such as computational power and data management software. In addition, a system of audit and control is required to evaluate the whole process.

5.1.2 Applying big data in service providing and policy making

The way in which the internal governance process is shaped is dependent on the goals that governments wish to achieve: adding public value. Public value is added through two specific tasks of the government: public service provision (services or products that are offered by the government and the public can make use of) and policy making (all processes involved with making policy, from the objective definition phase to the implementation and evaluation stage. A public service may be the result of policy).

5.1.2.1 Service provision

Governments can use big data in the services they provide to their citizens. Governments offer a variety of services to their citizens, such as transportation infrastructure, social services (e.g. public housing, social welfare), waste management, health care, military defence, safety etc. Typically these are goods categorized as public goods, i.e. goods that are non-excludable and non-rivalrous, goods of ‘public interest’, or goods that are considered to be underrepresented by the market.

Practical examples of how big data is used in public services are: the allocation of social housing by using an algorithm to determine who is in most need of accommodation; combating school drop-out by pre-emptively identifying students that are likely to not finish their education; pre-emptive surveillance in neighbourhoods with high crime rates; rating of teachers; identifying citizens that are likely to commit tax or social welfare fraud; detection of infections in infants; providing meteorological information and weather forecast; complaint processing of insurance claims, etc. (Bouma, 2016; O'Neil, 2016; TechAmerica, 2015).

Other examples are online data portals like Danish Borger.dk or Dutch DigiD. These are central places where citizens manage their data that is maintained by government authorities, for example for tax declarations or requesting rent allowance. Such portals collect and organize data from a variety of government departments (Giest 2017). Moreover governments collect and disseminate big data through sponsored statistical agencies such as the Dutch *Centraal Bureau voor Statistiek,* Germany's *Statistisches Bundesamt* or EU's Eurostat. These open data portals offer data on a whole range of topics such as transportation, economics, finance, demographics, labour market and more. The goal of these agencies is to collect and process data that can be used by the people, the media, businesses and science institutions for innovation and providing transparency (Bertot et al. 2014).

5.1.2.2 Policy making

Big data may be used by governmental institutions for policy making by broadening the evidence base or for predictive analysis. The policy cycle reflects the different stages of a policy life cycle from its start to the implementation and evaluation of the policy. The policy cycle has seven stages including a feedback cycle, indicating that the stages are non-linear.

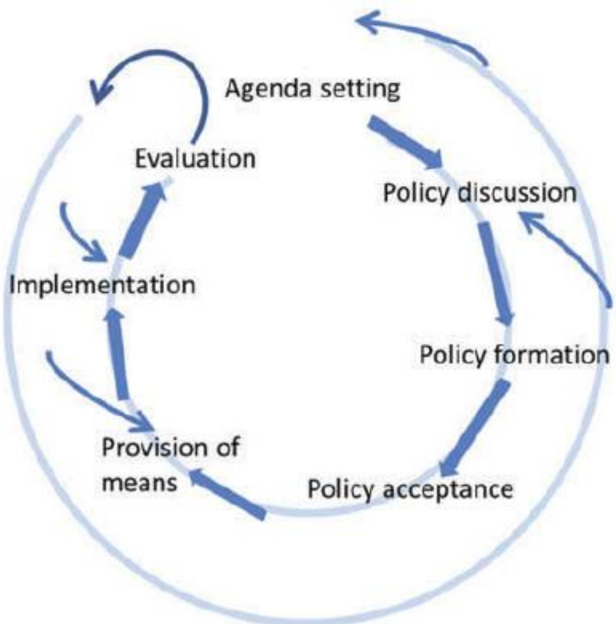

Figure 11 The policy cycle (Höchtl, Parycek, & Schölhammer, 2016)

Table 4 The application of BDA in the policy life cycle, based on (Höchtl, Parycek, and Schöllhammer 2016)

Stages	**Description**	**Role of big data**
Agenda setting	Identify problems and potential course of action	Identify emergent topics and policy preference of citizens, e.g. by analysing social media data
Policy discussion	Now the policy problem has been identified, the different options are debated	Data can be collected to see where policy action is most necessary, or data can be structured according to policy needs
Policy formation and acceptance	The steps of policy implementation are described and their formal (i.e. through voting) and informal (i.e. in terms of public opinion) acceptance are assessed	Higher predictive value adds legitimacy to policy by providing it with evidence, increasing its credibility and acceptance
Provision of means	Laying the groundwork for actual implementation, such as personnel acquisition and budget	Analysing earlier experiences to budget more efficiently, as well as providing more support for a policy if big data evidence shows it to be cost-effective
Implementation	The actual enacting of policy	Deciding on policy intensity, and since new data is being generated, it can be used in its continuing implementation.
Evaluation	Assessing enacted policy	Big data allows for real-time processing, which enables evaluation to be continuous and allow for course corrections of policy

The application of BDA in the policy life cycle is summarized in Table 4. In this way, Höchtl et al. (2016) write that big data enables a quicker decision making process, using more information and thus a broader evidence base, more inclusion of the public in the process, and a better usage of the abundance of unstructured data that is out there.

5.1.3 Challenges

Whether in public service provision or in policy making, there are various challenges that governments require to address regarding BDA. The following concerns are discussed here: the lack of skilled workers, the abundance of big data, government compartmentalisation, subjectivity of BDA, technocratic governance and corporate led governance. The issue of privacy is discussed separately due to its importance in chapter 7.

Skilled workers

One of the greatest challenges in using big data is to find the right personnel to manage and analyse big data, as the necessary expertise is neither simple nor only always technology oriented (Kim, Trimi, & Chung, 2014, p. 80). The skills required to find and analyse big data is complicated. Not only due to the required technical knowledge, but also because the translation of data to its practical application is not a straightforward task. Being able to describe patterns and make predictions are of little use when no practical issues are solved. Hence data analysis requires data scientists who, besides understanding the technology behind big data, have cross-sectoral expertise and understand real-world questions that are addressed through big data. In general, there is a lack of big data know-how within the public sector as big data literacy among civil servants and policy makers is often low (Margretts & Dunleavy, 2013). This makes finding enough skilled individuals one of the greatest challenges for governments for developing BDA (Kim, Trimi, and Chung 2014; Lillo 2018; Giest 2017)

Digital exhaust

Another challenge is optimizing the usage of all the data that is collected (Mergel, Rethemeyer, and Isett 2016; Milne and Watling 2018). People leave behind *digital exhaust* everywhere they go: location-based services register movement, speech activated devices recorded sound, surf behaviour and commenting on social media may reveal an underlying pattern, perhaps when combined with other types of data such as location, behaviour patterns or demographic information. A lot of data is collected without a specific purpose yet. Or data is collected by private parties that are useful for governments as well. Using as much data as possible is tempting, however, the standard BDA challenges related to the 5 V's need to be taken into account.

Government compartmentalisation

Another issue is the compartmentalisation of the government, where different government departments operate with their own methods of collecting and storing data, a reluctance to share the data and the absence of a legal framework which enables data sharing (Kim, Trimi, & Chung, 2014; Giest, 2017). The organisation structure of the government is best characterized as 'silo's', bringing difficulties regarding data exchange, standards, formats, technologies and so on. In addition, government agencies operate according to different legal frameworks that demand different ways of handling data, further complicating the sharing of data within the government. Data integration between different levels of government is required to unlock the full potential for big data with increased interdepartmental collaboration as many different parts of the government require each other's information.

BDA is biased

While there are many successful examples of BDA applications by the government, Giest (2017, p. 378) notes that "some have raised concern that policy decision-making based on the data mining models will exacerbate bias and create new forms of discrimination, due to algorithms that reflect norms and values and could reinforce structural inequalities and cumulative disadvantages". This is caused by something called 'algorithmic fairness' or 'algorithmic accountability', meaning that attention should be paid to how algorithms are constructed and how they operate. Algorithms are only as good as the scrutiny with which they are constructed. GIGO, or 'garbage in, garbage out' applies here, as algorithms may include inaccurate, wrong or irrelevant proxies. There seems to be a general misconception about the apparent objectivity of BDA (Graham and Shelton 2013). Rather, data and algorithms are not value neutral, they are inherently subjective due to their creator or social context in which BDA is embedded (Kitchin 2014; Graham and Shelton 2013). The complexity of BDA makes this increasingly problematic. Algorithms are difficult to understand if you are not a data scientist, evoking the feeling of being subject to a Kafka-esque system (Eubanks, 2018; Graham & Shelton, 2013).

Moreover, when decisions on public service provisions are taken over by algorithms, feelings of accountability for social obligations or empathy are disappearing. When people are excluded from public services or receive penalties, the difficult, though and non-evident decision underlying that outcome are ignored and replaced by an algorithm that is seemingly fair and objective. In particular when it involves disadvantaged people, who often lack the means to challenge the algorithms and who are dependent on public services (Eubanks, 2018; O'Neil, 2016).

Technocratic governance

Another concern is that big data can lead to a style of governance that becomes too heavily focused on addressing problems with big data rather than taking into account other factors that fail to be captured by big data. Kitchen (2014, p. 9) notes in city governance: "the drive towards managing and regulating the city via information and analytic systems promotes a technocratic mode of urban governance which presumes that all aspects of a city can be measured and monitored and treated as technical problems which can be addressed through technical solutions". "However" Kitchen continues, "technocratic forms of governance are highly narrow in scope and reductionist and functionalist in approach, based on a limited set of particular kinds of data and failing to take account of the wider effects of culture, politics, policy, governance and capital that shape city life and how it unfolds." In short, employing big data analysis evokes a potentially false sense of providing evidence, rationality, objectivity and accountability that endangers government action.

Corporate-led governance

With governments seeking to apply big data technologies, businesses supplying governance services see governments as their potential customers. Kitchen (2014) writing, on city governance, observes that "The smart city agenda and associated technologies are being heavily promoted by a number of the world's largest software services and hardware companies who view city governance as a large, long-term potential market for their products. Either through being major partners in building cities from the ground up (e.g., Songdo or Masdar City), or partnering with established cities to retrofit their infrastructure with digital technology and data solutions, these companies have been seeking to make their wares a core, indispensable part of how various aspects of city life are monitored and regulated." The results is that the ICT solutions is the effect of a push from sellers to governments to use big data rather than a demand from governments. The need for private technologies by government is confirmed by Geist (2017), who states that the lack of knowledge on big data is the main reason for outsourcing, rather than cost-efficiency. Governments requiring support from the private market is inevitable, however, it is important that governments remain in control of the products they need.

5.2 Governance in transport

Trends suggest that the usage of big data by government in transport operations is becoming mainstream. Big data collection and analysis technologies have significantly increased the availability of data on transport related phenomena. At the same time, there is a large increase in the availability of data with a component relevant for transportation, such as location tags in social media data, and transport data with no specific purpose of analysis yet, mobile phone data with a locational component (Milne and Watling 2018). To understand the role of governments regarding transport related BDA, in this section, the effect of BDA on the three transport related government tasks are considered: transportation planning, monitoring and surveillance of traffic, and providing public transport.

5.2.1 Big data in transport policy making and service provision

5.2.1.1 Transportation planning

Transportation planning refers to political and technical issues related to the planning and construction of transport infrastructure. It involves the process of policy making, gathering the required input for policy making and the challenges involved in realizing the transport infrastructure. In particular, BDA facilitates transport planning by collecting data on transport flows in support of the evidence base. Moreover, BDA allows for immediate monitoring of policy interventions.

The primary source for transportation planning is data on traffic flows. In addition to tracking current traffic flows, transportation planning is concerned with understanding the *cause* of current traffic and *predicting* future traffic. Understanding the origin of traffic is done by understanding the spatial and temporal distribution of activities of people. Hence, the scope of the data that is used to identify the timing, location and frequency of people's activities is much broader than just recording traffic flows. Potential data sources may be commercial transactions to identify where people shop at what time (and where restocking is required), social media data to identify popular locations, or GPS data from on-board computers. Other sources for understanding traffic next to human behaviour are the location and condition of infrastructure or geometrical information.

With the advent of big data, it is possible to expand the evidence base of transportation planning by collecting more real world evidence. Larger samples can be gathered by new and automated data collection methods that were previously done through surveys, for example through GPS tracking, sensors, online questionnaires or analysis of social media data. In addition, data can be collected from parameters that were previously difficult to record, such as the effect of parliamentary cycles. More data offers more insight in specific scenarios, such as weekly, daily or hourly patterns, influence of different weather conditions, effects of specific events, such as bridge opening, traffic jams, construction works and accidents, or impact of different demographic groups such as students, elderly or commuters. More data allows transportation planners to have a better understanding of the roots and causes that generate traffic flows and take into account specific conditions or events. Subsequently, BDA facilitates planners with more accurate predictions that allows them to plan transport accordingly. A rather unconventional way to use big data in transport planning is its application by the EU in military mobility, which further elaborated in Box 3.

Box 3: Using big data for Military Mobility

The use of Big Data could also benefit the transport and logistics in defence operations. Planning and route optimization improves efficiency within the logistic sector since logistics analytics can optimize inventory and resource allocation, identify optimal distribution locations, and minimize transportation costs (Sanders, 2016). Large amounts of military data are generated daily from intelligence, surveillance and reconnaissance (ISR) technologies. This data can be applied to improve situational awareness by, for example, deploying algorithms that could identify objects of interest (Miltner, 2018)). Although military operations are often complex, the concept of using Big Data analytics to provide (real-time) recommendations could be a decisive advantage.

As of 2017, the European commission is stepping up efforts to improve military mobility.[35] This has led to Action Plan on military mobility.[36] The aim of the Action Plan is to work towards a 'fully-fledged' European Defence Union. This must ensure smooth, efficient and effective movement of military personnel and assets across and beyond the EU. Cross-border mobility is momentarily hampered by a number of barriers that can lead to delays, higher costs and/or increased vulnerability.

The barriers relate to the unsuitability of infrastructure for the weight or size of military assets (insufficient height or load capacity of bridges/railroads) and lengthy and complex procedural issues that may differ per member state (custom formalities related to military operations and rules on the transport of dangerous goods in the military domain). A pilot exercise identified opportunities for the dual-use (civilian & military) of transport infrastructure. This included: multimodal platforms allowing to quickly shift assets from ports and airports to rail and road, improving the capacity of inland terminals and adequate loading gauges on freight rail lines.

One of the actions involves determining whether TENtec, EU's massive information system, can be used for planning military mobility in addition to its current usage of planning civilian infrastructure. TENtec contains all kinds of information on EU's transport network, such as the quality of route links and bridge heights. The database can be expanded to include parameters such as the vulnerability of bridges, i.e. information that aid in the planning of defence logistics.

While increasing the evidence base by collecting and analysing more data can be a costly and time consuming exercise, at the same time, it reduces the planning process by providing more flexibility to planners (Milne and Watling 2018). Continuous data collection and monitoring during the implementation phase allows for experimentation through a 'learn-by-doing' approach, reducing the initial research costs. Now the complexity of transport planning can be dealt with when multiple scenarios can be tested in real life. Constant monitoring allows planners to keep track of the effects, and small changes to the policy can be made if necessary. In Table 5 the BDA framework for policy making of Höchtl and his colleagues (2016) is applied to transport planning.

[35] European Commission (2017). JOINT COMMUNICATION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL - Improving Military Mobility in the European Union.

[36] European Commission (2018). JOINT COMMUNICATION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL - on the Action Plan on Military Mobility.

Table 5 The BDA framework for policy making applied to transportation, based on (Höchtl, Parycek, and Schöllhammer 2016)

	Description	**Role of big data**	**Role of big data in transportation**
Agenda setting	Identify problems and potential course of action	Identify emergent topics and policy preference of citizens, e.g. by analysing social media data	Analysis of infrastructure flows and the spread of human activities to identify bottlenecks
Policy discussion	Now the policy problem has been identified, the different options are debated	Data can be collected to see where policy action is most necessary, or data can be structured according to policy needs	Traffic forecasting for assessing the impact of various scenario's
Policy formation and acceptance	The steps of policy implementation are described and their formal (i.e. through voting) and informal (i.e. in terms of public opinion) acceptance are assessed	Higher predictive value adds legitimacy to policy by providing it with evidence, increasing its credibility and acceptance	Effects of transport interventions are supported by evidence, enabling a better justification of spending public resources and possibly higher public acceptance
Provision of means	Laying the groundwork for actual implementation, such as personnel acquisition and budget	Analysing earlier experiences to budget more efficiently, as well as providing more support for a policy if big data evidence shows it to be cost-effective	Analysis of earlier transportation projects in terms of budgeting and planning facilitates for better project preparation. Broader evidence base enables a more accurate calculation of required means
Implementation	The actual enacting of policy	Deciding on policy intensity, and since new data is being generated, it can be used in its continuing implementation.	Big data enables immediate tracing of policy effect by recording transport flows or other effects to maintain or adjust policy direction. This allows for a learning by doing approach
Evaluation	Assessing enacted policy	Big data allows for real-time processing, which enables evaluation to be continuous and allow for course corrections of policy	Implementation of transport planning can be immediately evaluated

5.2.1.2 Traffic monitoring

Currently, most big data technologies in transport are used for monitoring traffic (Hu and Liu 2017; Shi and Abdel-Aty 2015). Having insight in traffic flows is not only valuable input for transport planning, traffic monitoring is part of providing a service to society to ensure that traffic keeps flowing and avoid congestion. Big data is used by traffic management authorities for real-time analytics, control and supplying personalized information to transport users. The goal of traffic monitoring is to improve the efficiency of traffic operation, increase traffic safety, and reduce the impact of traffic on the environment (Hu and Liu 2017). Big data technologies are used to reduce congestion, predict incidents, estimate travel times and traffic simulation.

Data for traffic monitoring can be collected by three types of technologies: in-road technologies, over-roadway technologies, or off-road technologies (Martin, Feng, and Wang 2003; Shi and Abdel-Aty 2015). In-road detectors are placed in or across roads or pavements, such as inductive loops. Over-road detectors are places over or alongside roadways and record traffic from a distance such as speed or red light camera's. Off-road detectors typically refer to in-vehicle devices such as GPS location or Bluetooth technologies. Connected vehicles use plenty of off-road technologies, but the downside is that not every vehicle has these technologies. Data generated on the road can be connected and integrated to other data sources, such as demographic, geometric or meteorological data, as well as social media data or data from mobile phones.

Big data enables the reconstruction of the specific traffic conditions to identify causes of incidents or congestion. Shi & Abdel-Aty (2015) used 275 detectors to record traffic along 120 highway kilometres. By collecting real time data, the authors identified a relationship between congestion and the likelihood of rear-end crashes. This information was used to identify what the appropriate time was to raise speed warnings in relation to congestion intensity to prevent rear-end crashes. Real time tracking of traffic allows for taking into account the uniqueness of various traffic situations, such as various levels of congestion and weather condition. Another example is the usage of BDA to video material. Singh, Vishnu, & Mohan (2017) demonstrate how the large amounts of video data collected from road cameras can be used to identify traffic violators. The authors developed an automated system to process large amount of visual material to identify bike-riders without a helmet in traffic.

5.2.1.3 Public Transportation

Public transportation refers to the services provided for the transportation of passengers, usually according to a fixed schedule and along fixed route.

The use of transit passes enables public transport operators to keep track of the origin and destination of their passengers, which allows for identifying location and timing of travels made. Knowing where and when your passengers travel is crucial for an optimal allocation of the available rolling stock in way that best matches travel demand. Whereas such travel information was previously collected through surveys that only covers a small sample size, transit passes, enable data to be collected from entire population.

When origin and destination data is connected to other data sources, new patterns may be revealed. For example, rain can cause certain routes to become overcrowded so that additional transportation on these routes are required. Travel patterns of demographic groups can be used to develop specific products aimed persuading new customers to make use of public transportation. Collecting all kinds of information from their own rolling stock enables public transport planners to improve their service as well. Tracking the location of vehicles is used for providing real-time travel information to the public. Moreover, if certain routes show systematic delay, operators can investigate the causes.

Big data analysis is not exclusively being used to track the location of vehicles, goods or passengers. Liu & Yen (2016) show how big data can be used in automating the processing of complaints on bus passenger services to provide better services to the public.

5.2.2 Challenges

It is evident that there are many ways in which big data may be useful for governments to manage their transport operations. But to unlock its full potential, a variety of challenges remain.

5.2.2.1 Making use of all the data

Many big data technologies exist in the form of ITS, but the current generation of ITS suffers from single system functioning, lack of integration, lack of backward technology and the lack of universal standards (Zeng 2015; Hu and Liu 2017). The result is that much of the available data is not used sufficiently or used ineffectively. In general, there is a lack of utilizing the data generated by transportation, in particular for unstructured data (Wu and Chen 2017). The digital exhaust that people provide the government is not fully being exploited.

Exploiting the full potential of big data requires collaboration within the government and with the private sector. Other government departments record data on people's activities, such as data on commercial activity or other patterns of human behaviour that influence their travel patterns. The existing compartmentalisation of the government provides a challenge regarding the data exchange between different departments (Giest 2017; Kim, Trimi, and Chung 2014).

Moreover, data owned by private parties can be valuable for transport planners as well. Examples are collaboration with owners of location-based data such as smartphone providers, application developers or providers of navigation services. Likewise, car manufactures processing data of connected vehicles a variety of valuable input, for example for maintenance works when vehicles detect a bump in the road.

Currently the main barrier for public-private data exchange is the absence of a sufficient legislative framework that resolves issues of data ownership, privacy protection and appropriate compensation schemes. The European Automobile Manufacturers Association (ACEA) for example notices the societal value of data generated by vehicles and highlights the absence of a legal framework that addresses these issues, which makes car manufacturers hesitant to share their data (ACEA 2016b, 2016a). In addition, data owners – people and companies directly supplying data or disseminating data in the open sources format – may become aware of the value of their data and demand more compensation than currently is the case. At the same time, if data from private parties is used, policy makers become dependent on data supplied by private parties for policy making. This can be problematic when the input for policy making operates according to corporate interest rather than societal interest, or it can lead to a technical lock-in (Kitchin, 2014, p. 10).

5.2.2.2 Limits of big data

With the potential ascribed to big data, transport planners can be subject to greater expectations (Milne and Watling 2018). In particular, there seems to be an misunderstanding about the potential of big data, as big data is used to answer questions that exceed the scope of what big data can do (Graham and Shelton 2013; Kitchin 2014). Issues arise when big data is used to answer questions about processes that are too complex to be datified.

Another issue relates to the ethical[37] dimension of transport planning. Big data may reveal that that certain demographic groups behave differently. What if it turns out that certain groups are more prone to cause delays or traffic accidents than others? This can lead to discrimination policies, for example in the commercial insurance market (Milne & Watling, 2018, p. 8).

5.2.2.3 Big data methods

In addition, Milne & Watling (2018) describe a more fundamental issue to how big data has been applied in transport planning. So far, big data has been mostly used to replace 'small' data in existing analytical approaches and models. However, the authors argue that potential of big data in planning transport is larger than expanding existing approaches of policymaking and can be used in new ways to discover new things about transport systems. They argue, planners and academics "need to engage in a fundamental reassessment of what data can tell us about transport systems that help us better understand how they function and what we can do to influence them in positive ways" (Milne & Watling, 2018, p.8)

5.3 Summary

Similar to how businesses use big data to draw abstractions and make predictions to increase efficiency and profits, government exploit big data to achieve their goals. This includes applying big data for a more efficient government apparatus, a broader evidence base for policy making, or providing better services to their citizens, e.g. a better distribution of social services or crime and abuse prevention due to early signalling of potential offenders. With respect to the government's role in managing transport operations, BDA can be used to improve traffic monitoring, public service provision or transport planning.

Governments applying big data do not have to start from scratch – many governments, academia and the private sector are using big data. Giest (2017) notes that the government has moved past the question 'if' big data is to be used. Nowadays it is about the 'how' big data can be incorporated in the governance process and 'how' it can be utilized in public service provision and policy making. Regarding the 'how' questions, Giest (2017) identifies two main challenges: "First, the fact that existing administrative and institutional structures define the way data is collected, analysed and used due to limited institutional support and data silos. And

[37] D 2.3 of WP2 deals specifically with the ethical dimension of big data.

second, the capacity within government plays a role in how data is dealt with or used at all when looking at specific policy domains."

Big data can improve governance by the government, but there are some concerns. Applying big data is challenging due to its complexity and due to the misconceptions surrounding its potential. In particular for governments, the lack of skilled workers and the compartmentalisation of the government departments makes exploiting big data in a way that adds value challenging. Moreover, there are some limitation to the potential of big data. Subjectivity can creep into presumed objective algorithms and not every social process can be datified. Applying big data in the government process, either in the government organisation, in public service provision or in policy making requires careful planning, and there are various roadmaps available to ensure a meaningful integration in government governance, one of which is described in this chapter.

Most of the general issues regarding the application of big data by governments are observed for transport related issues as well. There appears to be a lot of unexploited data, in particular in possession of private parties. Challenges exists in making use of all the data and developing a legal framework to do so. In addition, there is plenty of unlocked potential for governments to use BDA to answer new questions in new ways, rather than answering existing questions in faster ways. While there are no signals of wrong application of BDA, transport planners need to consider BDA methods while facing high expectations of BDA.

6 International Level: Big Data in International Relations

The aim of this chapter is to uncover the international dimension of big data by examining the issues involved when data crosses the border. We consider the 'political pressure points influencing the flows of data across the border'. In order words, this chapter describes *how* cross border data flows are impacted and *why,* with the aim to identify the issues involved when data crosses the border.

Another part of international relations literature is concerned with how big data influences diplomacy, the role of big data in national security, or the geopolitics of big data. Due to the limited relevancy of these topics to transportation, they are not further elaborated here.

6.1 How cross-border data flows are regulated

Data moves through cyberspace i.e. the notional environment in which communication over computer networks occurs[38]. Cyberspace is an "complex technological environment that spans numerous industries, governments and regions" (Deibert 2015, 13). There is no one nation or institution managing cyberspace, hence, "governance is spread throughout numerous small regimes, standard setting forums, and technical organizations from the regional to the global" (Deibert 2015). Governments can impose restrictions on what can be done with data, for example by conditioning data flows (e.g. privacy regulations, such as EU's GDPR) or restricting storage and processing location of data (which is often the case for classified information or financial information). Such restrictions limit the flow of data through cyberspace and restrict data usage of data suppliers and data users. Since 1990 legislation on cross-border flows of data has proliferated, see Figure 12.

Limitations on cross border data flows are imposed by policymakers to maintain control over data that may leave a country. In Romania, data on Romanian online gambling users and the game server itself requires to be located in Romania[39]. In Luxembourg, the IT-infrastructure for financial data, particularly data that is under supervision of national regulators, can only be outsourced when strict preconditions are fulfilled (Graux, Ypma, and Foley 2017, 129).

Regulating cross-border data flows is done by imposing conditional measures for data transfer or by either restricting the movement of data. Restrictions may be imposed on the storage location, processing location or data transfers (Ferracane 2017)

- **Storage location** restrictions require the data to be stored within the country where the legislation applies. Data storage and processing may happen abroad, as long as copy of the data is stored domestically. Types of data to which this requirement applies is often corporate or social documents such as tax records, accounting documents or financial reports.

[38] https://en.oxforddictionaries.com/definition/cyberspace

[39] http://www.rombet.com/uploads/legislation/Norms%20HG%20111%20Gambling%20law%20Romania%20ENG%2026_02_2016.pdf

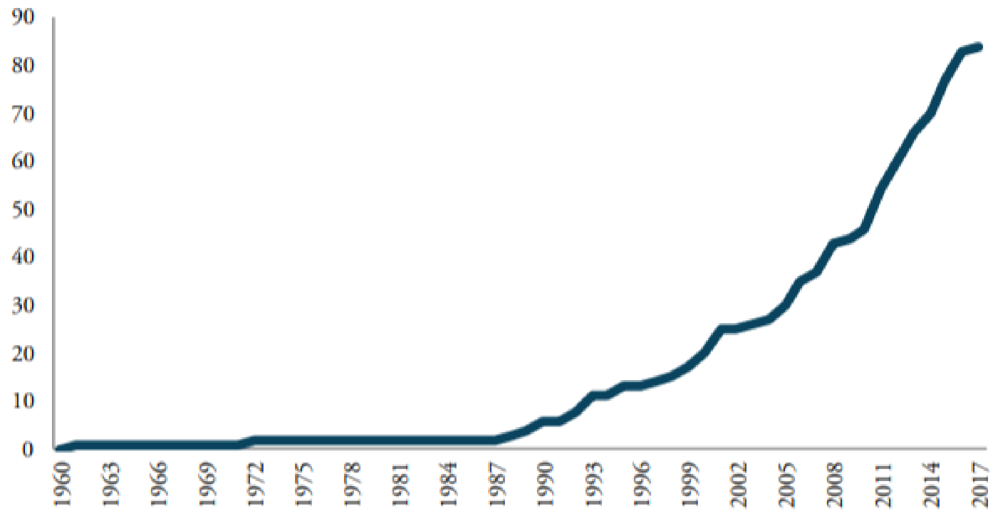

Figure 12 Number of restrictions on cross-border data flows (vertical axis) between 1960 and 2017, based on 64 countries. Source: (Ferracane 2017)

- Restrictions on **storage processing** demand that the main processing of data is taking place in the country that the regulation applies to. A local data centre or local provider is required. After the main processing, the data may still be transferred abroad for minor manipulations. Processing constraints are typically applicable to the processing of data on a nation's citizens.
- Most severe are restrictions on **data transfer**, which require the data storage, processing and usage to be within the same country. In this situation, sending a copy of the data abroad is prohibited. There restrictions are often applicable to certain datasets, such as public information held by government institutions, health data or financial data.
- In addition to restrictions, **conditional measures** enable the cross-border flow of data but under specific terms. An example is the GDPR, where data transfer outside the European Economic Area (EEA) is only allowed if the receiving institution has an adequate level of privacy protection. Infrastructure requirements may also be conditional, if companies require to be in possession of a local server to handle the data.

Graux, Ypma, & Foley (2017) researched data policies in Europe and distinguishes between policies imposing direct and indirect barriers. Direct barriers refer to obligations by legislation, e.g. when data needs to be stored in a specific country. Indirect barriers refer to laws that likely impact data storage and flows, e.g. data may not be given to third parties or supervisors or regulators require data access. The type of barriers are described in Figure 13.

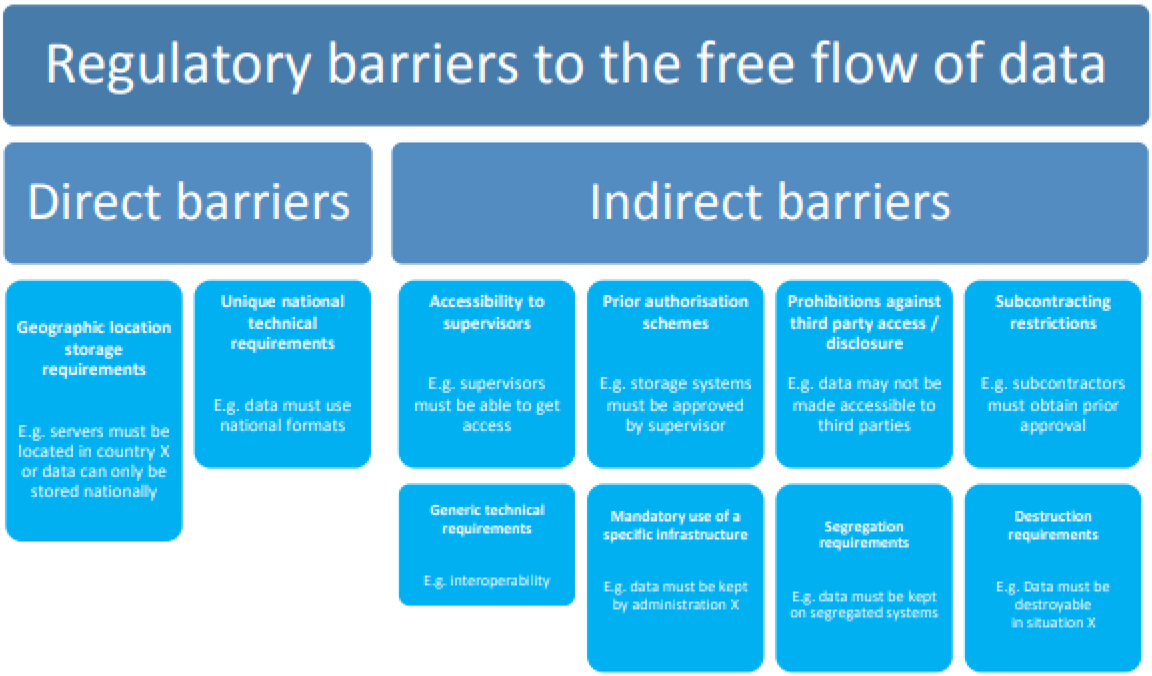

Figure 13 Direct and indirect regulatory barriers to the free flow of data existing in Europe. Source: (Graux, Ypma, and Foley 2017)

Considering these measures, there is no clear relationship between the regulation and the goal of regulation (Graux, Ypma, and Foley 2017). Requiring particular data to be stored locally can have multiple reasons, such as protecting privacy, limiting potential breaches or imposing commercial restrictions on foreign companies. To fully understand the political pressure points of big data, an examination of the motivations behind cross-border data policies are required.

6.2 Motivations regulating cross-border flows

Reasons for limiting data are often privacy or security related, as the less data is available, the smaller are the chances of misconduct or data breaches (Graux, Ypma, and Foley 2017; Ferracane 2017). In the Netherlands, public sector data of a classified nature needs to remain in the Netherlands as abroad storage is conceived to involve risks that cannot be covered yet[40]. Another reason is that geographical containment of data makes it easier for regulators or supervisors to keep oversight (Graux, Ypma, and Foley 2017). It significantly eases the job of national inspection authorities when accounting documents or financial reports are stored domestically. Other reasons for limiting dataflow is when it leads to societal harm, i.e. theft, intellectual property right infringement, or for achieving more controversial aims such as limiting political freedom or employing commercial restrictions (Meltzer 2015).

[40] https://zoek.officielebekendmakingen.nl/ah-354118.pdf

6.3 Effects of regulating cross-border flows

While governments seek to regulate data flows for the public good, imposing restrictions can have negative consequences for data-based innovation. Generally speaking, requirements to data flows through cyberspace limits organisations in their usage of the data. The potentially destructive effect of cross-border data flows is reinforced by commentators, who question whether current legislation is appropriate. This is in part because “the novelty of the data revolution and the difficulty of policymakers to grasp its transformational impact on the economy led to responses that impose significant costs on the economy” (Ferracane 2017). There is potentially a lot to be gained from alleviating cross-border data restrictions, because the underlying motivations (such as privacy protection, national security, cyber security, law enforcement) can still be achieved with much less restrictions (Ferracane 2017; Graux, Ypma, and Foley 2017).

Another issue is that companies impose stricter cross-border policies than is necessary by law due to a perceived requirement of these policies (Godel et al. 2016). Companies employ stricter data residency requirement that do not have any foundation in official legislation, which is the result of a misinterpretation of the legal requirements.

Currently, the EU is addressing the barriers to cross-border data flows as one of the pillars under EU’s road towards ‘the digital single market’[41]. The EU considers the free movement of data as an additional EU freedom, in addition to the free movement of goods, services, capital and people. In 2018 the GDPR went into force regarding the free cross-border flow of personal data, combining the different legislations of the EU-member states into one. The GDPR has already been discussed at length in D1.2 on ‘Big Data Policies’[42].

More revolutionary are the efforts by the EU to work towards the free flow of non-personal data. In June 2018, the EU reached a provisional political agreement on the free flow on non-personal data[43]. Similar to the GDPR, legislation on free flow of non-personal data seeks to abolish the plethora of national legislations currently existing. The motivation was to deal with the following main obstacles (European Commission 2017):

- Data localisation restrictions
- Obstacles to data movement across IT systems
- Legal uncertainty, in part because of the complexity caused by many national regulations
- Lack of trust due to security risks and concerns about the cross-border availability of data for regulatory purposes

The new legislation contains the following components (European Commission 2017):

[41] https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy

[42] Accessible via: https://lemo-h2020.eu/newsroom/2018/5/31/deliverable-12-big-data-policies

[43] https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-parliament-and-council-framework-free-flow-non-personal-data

- The **free flow of non-personal data principle** removes unjustified data localisation restrictions imposed by public authorities, enhancing legal certainty and raising trust.
- The principle of **data availability for competent authorities** makes sure that the data remains accessible for regulatory and supervisory control also when stored or processed across borders in the EU.
- Actions to encourage cloud service providers to develop self-regulatory **codes of conduct** for easier switching of provider and porting data back to in-house servers.
- **Security requirements on data storage and processing remain applicable**, also when businesses store or process data in another Member State. The same applies when they outsource data processing to cloud service providers.
- **Single points of contact** in each Member State, to liaise with other Member States' contact points and the Commission to ensure the effective application of the new rules on the free flow of non-personal data.

6.4 Summary

The number of regulations on conditioning the cross-border flow of data has increased significantly. Such regulations fall into four categories:

- **Storage location** restrictions require the data to be stored within the country where the legislation applies
- Restrictions on **storage processing** demand that the main processing of data is taking place in the country that the regulation applies to
- **Data transfer** restrictions require the data storage, processing and usage to be within the same country.
- **Conditional measures** enable the cross-border flow of data but under specific terms.

Reasons for limiting data are often privacy or security related, or the geographical containment of data makes it easier for regulators or supervisors to keep oversight. Other reasons are more controversial, such as limiting political freedom or commercial restrictions.

While governments seek to regulate data flows for the public good, imposing restrictions can have negative consequences for data-based innovation. Commentators note that current cross-border data legislation is too destructive for innovation, and the same aims can be achieved with less harsh restrictions on cross-border data flow (Graux, Ypma, and Foley 2017; Meltzer 2015; Ferracane 2017) – similar to how the WTO addresses international barriers to trade. However, an equivalent of the WTO for international data flows is not existing.

The EU is currently addressing inefficiencies related to cross-border data flows. While the GDPR is well known, less known is that in June 2018, the EU reached a provisional political agreement on the free flow on non-personal data.

7 Privacy Issues of Big Data

The issue of privacy has been one of the most discussed social consequence of big data (Graham & Shelton, 2013, p. 258). This chapter discusses these issues, on a general level and specifically for the transportation sector. At the end, potential solutions for these issues are presented.

7.1 General issues

Big data often entails personal information such as age, address, zip-code, passwords, academic background, geographic location, professional background, opinions, interests, preferences, health, electricity usage and so on – any kind of information that can lead to the identification of a person. What classifies as personal information is still much debated as it can refer to any factual or subjective information about an individual (Yao, Rice, & Wallis, 2007; Rubinstein, 2013). What is considered to be private information differs from person to person and changes per context. This not only concerns data that is directly relatable to individuals. Meta data and the triangulation of multiple sources also enables the construction of detailed pictures of individuals. For example, using metadata on human mobility, Montjoye et al. (2013) were able to identify 95% of the individuals based on four spatial-temporal data points.

7.1.1 The value of personal data

Personal data is widely available with the emergence of data driven business models, in which people 'pay' with their personal data instead of money. Cookies trace user's path when surfing the web, registering what is looked at and for how long. Many websites or smartphone applications are free as long as the developer is given access to your personal information. Mobile applications extract more user data than is required for running the application by requesting access to geolocation, other installed applications, photos, messages, music and more. By now there are many business models in which (personal) data is a sufficient means of payment to receive a service for free or for a discounted price. Business are willing to collect data in favor of direct income as it yields more revenue in the long run, either by selling the data for advertisements, through product improvement leading to more sales or by selling user data to third parties that are willing to pay for it.

In general, people are not aware of how much their data is worth (Malgieri and Custers 2018). While people feel that their data is worth something, the value of data is difficult to determine, it is not easy to keep track of all your data and there is no market place available in which you can trade your data. (Personal) data is a valuable economic resource, but compensation schemes are largely absent. People are willing to give away their data for almost nothing, in particular when free services are to be obtained – even though people want to have more privacy (i.e. the 'privacy paradox'). Currently data is mostly 'used' instead of being owned and traded.

Various initiatives exists that enable people to monetize their data, such as CitizenMe or Brave. The former is an app that collects personal information which one can offer for sale. The latter is a browser that blocks online advertisements except for those that share their income with the data users. However, such user controlled data-selling schemes do not seem to gain popularity.

Above all, if companies pay their users for their data, it is the question if they can keep offering their services free. And even if users sell their data for market value, estimates suggest that on an individual level, prices are low. It is estimated that if Facebook distributes their profit among its monthly users, each will receive only 9 dollars a year (The Economist 2018). The value of general information (age, gender and location) is estimated at 0,05 cent per person; information on someone shopping for a car is 0,21 cent, data on a pregnant woman is 11 cent and personal health information is valued at 26 cent (Malgieri and Custers 2018).

However, in some instances, user data is more valuable. For example, self-driving cars are equipped with many cameras and sensors that keep track of the massive amounts of traffic situations that could occur while driving on the road. During each trip, a massive amount of data is collected on the interaction between the system, the driver and the environment; data which is crucial for the further development of autonomous vehicles, a field that is still in its infancy. In addition, the advent of AI and machine learning technologies increases the complexity of big data analysis and requires high quality data to be fed into the system (The Economist 2018).

But other than its monetary value, there are other risks involved which are difficult to capture in monetary value but make users hesitant to share their data. The danger is that personal data may fall in the hands of people or groups with malicious intent, or that the data is used for further purposes that are not compatible with the purpose for which the data was initially processed.

7.1.2 Why privacy is a concern

Individuals demand privacy for different reasons. Generally speaking, people are willing to sacrifice privacy if they can benefit from it or of they believe otherwise they are in danger. The level of privacy demanded by people depends on (1) the type of personal data that is being collected; (2) user's opinion on the technology for which the data is collected; (3) the amount of trust in the organisation that collects the data; (4) the type of organisation that is collecting the data; (5) the purpose of the technology for which data is used; and (6) whether the data may be shared or re-used after its initial use (Lederman, Taylor, & Garrett, 2016). The extent to which each of these six aspect influences an individual's choice to share data differs per person.

While each person can decide for their own if they want to share their data or not, there are more fundamental worries about why we want to limit the spread of personal data. In cases where permission is given to organisations to use personal data, for example through accepting the terms and conditions for using a particular service, there has been much debate whether users are 'actually' giving permission – usually they do not know to whom and for what they are giving permission. Concerns involved with the wide spread of personal data relates to the potential power held by big data owners when big data is used for predictive analysis and the related security issues. In this respect, the government can play a role in setting limits to what data is collected, or how it is used and stored.

7.1.3 Big data and predictive analysis

One of the main purposes of big data is to predict the behaviour of individuals so that services and products can be offered that better fit the needs of the customers. The predictions are based on individuals past behavioural patterns or on the choices made by people with similar behaviour or background. E-commerce companies offer product suggestions based on user history, product popularity and current trends. News websites offer reading suggestions to their readers based on algorithms that predict the interests of their users. Big data analysis reveals user preferences and enables the offering of products and services that more closely corresponds to the interests of the user. While making assumptions about individual's behaviour may be beneficial in certain situations, there are various reasons why there should be limitations to the assumptions that can be made about people. Assumptions about people's behaviour can lead to discrimination, exclusion and loss of autonomy (Kerr & Earle, 2013)

While some predictions are aimed at providing more and better options to the users, other predictions are more aimed at either limiting the users' options or steering behaviour in a certain direction (Kerr & Earle, 2013). These so called 'pre-emptive predictions' are in the interest of the maker to control behaviour rather than in the interest of the user. Pre-emptive predictions have been used in political campaigns to steer the opinion of people in such a way that they are more likely to vote for a particular candidate (Issenberg, 2012). Big data analysis is used to identify potential voters based on certain background and behavioural patterns, after which they are targeted with specific information in order to change their political views in the favour of a particular candidate. Similarly, big data analysis has been used to construct no-fly lists for individuals that are likely to be terrorists, or to determine the degree of supervision that is required for individuals on parole (Kerr & Earle, 2013). Students that are likely to drop out of school are identified based on big data analytics and provided with extra supervision (Bouma, 2016).

The concern is that with pre-emptive predictions, the agency of individuals is impacted. Those that are categorized to belong to a certain group or those who are conceived to show unwanted behaviour – for whatever underlying reason – are given special treatment before the behaviour is even acted out. Such steering techniques are often described by policymaker as 'nudging', which is the steering of behaviour of people for their own good. While there is high consensus that nudging is desirable in certain situations, such as discouraging smoking or unhealthy eating, there are many situations in which it is not evident (Sunstein, 2015) or where big data usage has led to discrimination and exclusion – even if it's often unintended (Eubanks, 2018; O'Neil, 2016). Eubanks (2018) describes how BDA led to a several disabled persons losing vital welfare assistance due to a computer mistake, or how homeless individuals which are in need of housing are systematically neglected because they do not conform to certain parameters that are included in the algorithm.

Some use big data to target vulnerable individuals with malicious intent, and there have been many cases where big data analysis has been carried out with bad and irresponsibly. And there are more controversial topics such as marketers using big data to identify the moments in which potential buyers are most vulnerable. Limiting the spread of personal information limits these negative consequences of big data.

7.1.4 Big data and security

Another concern relates to the issues of security. Generally speaking, when ICT systems become more complex, the weaknesses increases. Datasets become bigger by more data collection points (cookies, cameras, sensors, surveys and so on) and more database relations by more interrelated dataset connections. But the bigger the data, the more potential. A weakness in one of the many data collection points can expose the whole system. Last years witnessed the breach of multiple large datasets. In 2017, a breach in Equifax, a large credit reporting agency, exposed the data of 143 million people. A hack of cloud storage firm Dropbox in 2012 led to the usernames and passwords of 68 million users being stolen. A data breach in internet service company Yahoo in 2015 affected 3 billion users. Often, data protection has proven to be insufficient. There are insufficient guarantees that data is safe from coming into the hands of individuals or groups with malicious intent and fall victim to identity theft and credit card fraud.

7.1.5 Dealing with privacy concerns

In this regard, the government can play a role in protecting its citizens against the negative consequences of big data through legislation, either by limiting the personal data that is collected or by setting boundaries to how data is used. But governments need to be critical of itself as well in its usage of big data. The government have been responsible for bad algorithms and pre-emptive predictions (Eubanks, 2018; O'Neil, 2016) or weak security[44] as well. Having clear privacy legislation signals transparency and helps building confidence with costumers to share the data. With less consumer data, the products companies offer are of lower quality. If insufficient action is undertaken to protect privacy, it can result in an acceleration in voter demand for strict privacy legislation, such as the recent Facebook and Cambridge Analytics scandal has shown.

One way forward to ensure that companies and government can reap the benefits of the information that is revealed by collecting personal information while maintaining privacy of individuals is a shift in a data collection model towards ‘consumer empowerment’(Rubinstein 2013). This approach implies a shift “from a world where organizations gather, collect and use information about their customers for their own purposes, to one where individuals manage their own information for their own purposes—and share some of this information with providers for joint benefits” (Rubinstein, 2013, p. 8). Consumer empowerment can be provided by Public Data Services (PDS) – a third party that operates in the interest of consumers and provides them with a tool to keep track of their information and enables them give permission to companies to use their information for what purpose and for how long. PDS requires high level of security and user control, for example by tagging each piece of user information with metadata on the privacy related requirements (Rubinstein 2013).

[44] https://www.heritage.org/cybersecurity/report/federal-cyber-breaches-2017
and https://www.zeit.de/digital/datenschutz/2018-03/hackerangriff-bundesregierung-cyberattacke-angriff-hacker-russland-spionage

7.2 Transport specific issues

Privacy sensitive information in transportation involves the past, present and future location of individuals in relation to any other personal information. This section deals primarily with passenger traffic, as freight traffic concerns the transportation of goods which generally does not involve privacy issues[45]. In particular, the advent of Intelligent Transport Systems (ITS) can pose a threat for privacy, as it is based on the collection of large amounts of data. Below is a list of various types of personal data that can be collected in transportation, based Fries et al (2012) and Lederman et al (2016).

- **Red-light cameras** identify vehicles that run through red traffic lights and links the location and license plate of a vehicle with a vehicle ownership registry
- **Speed-detectors** measure the speed of vehicles in combination with the time and location
- **Traffic monitors** such as road sensors or cameras measure the general traffic condition at a road junction or road link to identify accident or bottlenecks. Increasingly, traffic monitoring cameras have become of higher quality, making it possible to distinguish individual vehicles or even the persons inside the vehicles.
- **Automatic tolling devices** record vehicles passing through a particular toll location by a transponder located in the vehicle or by camera's recording the license plate. A fee is automatically charged from the user's account.
- **Parking meters** record the amount of time a user is parked at a certain location. Increasingly, payments are done electronically, so that parking information can be connected to bank account information.
- **GPS tracking** are used in vehicles to keep track of its location, for example for navigation purposes, fleet tracking, parking information services or for governments to keep track of vehicles carrying hazardous goods.
- **Transit passes** are increasingly more common to use when travelling, in particular for public transport. The routes of users can be traced by linking the check-in point to the check-out point. While some are anonymous, other passes are connected directly to a payment account or other types of personal information.
- **Connected vehicles** are the most recent form of ITS, referring to vehicles being equipped with internet access, allowing a constant connection to other devices within and outside the vehicle. The potential for connected vehicles as a data source is high because the type and timeliness of the information that is provided is new.

7.2.1 Examples of various privacy issues

7.2.1.1 Vulnerability of transit passes

A general problem in all modes of transportation is the vulnerability of e-tickets or transit passes, i.e. tickets that allow the storage of multiple or unlimited pre-purchased trips. These electronic proofs of payment offer many advantages to both passenger and transport operator due to the flexibility, ease of use and the information generated.

[45] it may when it concerns the location of the driver on-duty, however, this is not considered here

Such tickets are typically based on Radio Frequency Identification (RFID) technology. But the security of these cards are not always guaranteed. In 2008, researchers were able to hack in and read the data from the MIFARE classic card - the most commonly used model for transit passes – with relatively basic tools (Gans and Hoepman 2008). The card is based on a relatively simple design due to low cost requirements and speed in usage. It is often a trade-off between enhanced security and enhanced efficiency.

Other ways of misusing transit cards are cloning or swapping. Cloning occurs when digital data stored on an electronic ticket is copied and transferred onto another device. This is done with considerable ease to a relatively unencrypted system (Sadeghi, Visconti, and Wachsmann 2008). Alternatively, the identities of two tags can be switched in a process known as swapping. E-ticket technology can be tapped into or 'eavesdropped' relatively easily (Sadeghi, Visconti, and Wachsmann 2008).

Personal data on the user of transit passes are not stored on the card itself but on related databases. The personal data may be spread when the company is not aware what personal information can be inferred from the travel databases. The Japanese company East Japan Railway intended to sell data collected by its ticketing system (such as travel history) to a third party, Hitachi. Although the company claimed that personal data had been anonymized, the level of privacy protection was questioned[46]. A surprising amount of detailed information about users can be inferred from something as simple as travel history. Information such as living patterns and likely places of residence and work can be inferred from the (at face value) anonymous data.

(Avoine et al. 2014) discuss the case of the bus system of Cesena, Italy, which collects data through its electronic ticket system MiMuovo. Even when anonymized, the behaviour of passengers reveals information about them through this tracking system. Cross-referenced with other publicly available data, it is possible to easily obtain the identities of passengers. In general, data from transit passes are used beyond simple analysis as frequently unique ID's are given to passes so that individual movements can be traced and linked to their identities (Milutinovic et al. 2015)

7.2.1.2 Privacy and ITS

The advent of ITS, i.e. road monitoring using smart technologies, brings challenges to privacy as well. Through gathering information on driver, ITS is used to tackle issues such as congestion and pollution. The system is also, however, criticized for violating the privacy of motorists (Cruickshanks and Waterson 2012). It is considered a 'major challenge' for big data companies to maintain the privacy of individuals due to the sheer volume of data produced by ITS systems (Gohar, Muzammal, and Ur Rahman 2018).

(Hommes and Holmner 2013) discuss the 2007 case of the South African roads agency (SANRAL) putting in place a number of ITS innovations on the freeway around the city of Gauteng. Variable message sign boards, CCTV and e-tags were introduced, to tackle issues such as congestion by

[46] See: https://www.japantimes.co.jp/opinion/2013/08/03/editorials/jr-sells-commuters-data/

charging a fee to motorists using a specific section of road during peak hours. According to the authors, the system raises privacy concerns as travel habits, vehicle information and even bank details are communicated through SANRAL's database(Hommes and Holmner 2013).

Other types of data-collecting technologies register driving behaviour from within the vehicle. Examples are section control (basing speed on time travelled across a certain section of road), ISA (intelligent speed adaptation) and EDR (event data recorder). The second also monitors the speed driven (relative to the allowed speed limit on a particular stretch of road) and warns the driver. The latter registers driving behaviour including braking and throttling, to be used as a 'black box' after an accident has occurred (Eriksson and Bjørnskau 2012). These systems make it possible for a large amount of data relating to driving behaviour to be stored and possibly used by other parties.

Connected vehicles pose another privacy question. This technology links vehicles together, and links them to a transport grid. Again, the collected data can also be used to monitor the travel habits of users, sparking privacy concerns. A cyber-attack on the network that links these cars could also compromise the privacy of drivers. In some instances, data sharing can be advantageous to the user, if such data can be shared with insurance companies it could lead to reduced premiums (Zaidi and Rajarajan 2015).

Other types of road travel adds new dimensions to the privacy question as well. Ridesharing, i.e. services that facilitate the sharing of car journey, is an example of a big-data oriented type of service where privacy plays a role. (Aïvodji et al. 2016) describe a particular issue of ridesharing related to the designation of meeting points by the system. 'Points of interest', or geographical locations favoured by users can be combined with additional data to extract information about the user and extrapolate their future movements. In their work, the researchers propose an alternative method to identify point of interests in which privacy of users is secured, while the quality of the ridesharing service is maintained (Aïvodji et al. 2016).

To facilitate digital between vehicles and between vehicles and road infrastructure, the EU adopted the European Strategy on Cooperative Intelligent Transport Systems (C-ITS) in 2016. The aim of the C-ITS strategy is to support investments in C-ITS, enable access to European funding for projects and to work towards an appropriate legal framework.

7.2.1.3 Privacy vis-a-vis security

The demand for privacy is not always the most pressing issue on the minds of passengers. (Potoglou et al. 2010) point to the trade-off between security and privacy in the railway sector. The need for increased security in this travel mode has been particularly stimulated by fears of terror attacks. The authors find that generally citizens are willing to trade off some privacy (e.g. camera recordings) for increased security, despite concerns about privacy (Potoglou et al. 2010).

After the September 11 attacks of 2001 air travel has been one of the modalities most obviously impacted by security regulations, some of which significantly impact the privacy of travellers. An example is the implementation of biometrics in identity management. Data relating to physical characteristics such as irises and fingerprints are included in passports and scanned at the airport. Combined with Passenger Name Record (PNR) data, these could form a major risk to air travellers' privacy. While such information is purely used for security and not, say, for innovation or product improvement, the degree to which privacy information is obtained is troubling. (Nouskalis 2011) points to the excessive violation of privacy the use of this data would have, especially in peacetime. Indeed, the combined biometric and travel data could reveal much more than simply the passenger's identity: it could even reveal health information (Crawford and Schultz 2014).

The predictive use of data is another issue that derives from the increasing amounts of data collected on each individual. A profile is created based on gathered information. This can have unintended consequences. The 'No Fly List' used by the American Transportation Security Administration is an example of this which is discussed by Crawford. This list is generated by big data and has the side-effect of mislabelling a large amount of people as terrorists based on profile creation (Crawford and Schultz 2014).

7.2.2 Dealing with privacy concerns

All big data applications contribute to increased transport safety, security and efficiency, but also involve concerns about privacy. Location-based data is collected and combined to other types of personal information such as user information and account details. Once the data is collected, it is stored and connected to other databases. For tolling information car images are connected to billing accounts to charge the right fee from the right user. Transit passes are connected to user ID's to record the routes that passengers travel. An UK survey held by (Cruickshanks & Waterson, 2012) reveals that some people (in particular elderly, poorly educated, females, those belonging to an ethnic minority, and people not regarded as tech savvy) are less willing to travel if it means giving up personal information. Consequently, more ITS can results in a reduced mobility of society as a whole.

Such concerns about privacy should be incorporated in the development of ITC systems. (Lederman, Taylor, and Garrett 2016, 1) propose privacy concerns to be addressed as follows: "through both **privacy-by- design** solutions (that build privacy protections into data collection systems), and **privacy-by-policy** solutions (that provide guidelines for data collection and treatment) including limiting the scope of data collection and use, assuring confidentially of data storage, and other ways to build trust and foster consumer consent"

Another way to ensure data privacy is through anonymization and pseudonymisation techniques. Various strategies exist while maintaining the data's usefulness for analysing transport operations. Aïvodji et al (2016) provide a list of techniques for anonymizing transportation data: aggregation, anonymity, cloaking, encryption, geographical masking, mix zone and pseudonyms.

- **Aggregation**: individual mobility is combined by time and space to prevent the identification of single routes
- **Anonymity**: K-anonymity refers to the level of anonymity in datasets and is achieved when each identity characteristic appears in at least k-amount of rows.
- **Cloaking**: a route is transformed in a similar route with lower spatial resolution
- **Encryption**: data is encrypted before being processed

- **Geographical masking**: exact locations are masked, for example by not recording the first and last kilometre of a trip
- **Mix zone:** routes are transformed from lines into zones
- **Pseudonyms**: user ID's are replaced with a pseudonym

Other methods are using thrusted third parties to manage the data, asking the individuals for permission to track them, or having legislation in place to set the privacy rules (Fries, Gahrooei, Chowdhury, & Conway, 2012). If transportation users have more faith in companies or government maintaining their privacy it is more likely that the users are willing to share their data.

7.3 Summary

The growth of data-driven business models (where consumers pay with their information for services instead of money) and the proliferation of data collecting techniques (such as cookies or advanced cameras) has significantly contributed to the processing of personal data. The increasing availability of personal data is worrying because it significantly increases the chance of privacy infringement. Ultimately, personal data can come into the hands of groups or individuals without the data subject being informed about it, and thus unable to exercise his/her rights recognised in the applicable legislation. Another issue relates to the predictive capacity of big data, which causes loss of autonomy and essentially turns around our notion of 'innocent until proven guilty' to 'guilty until proven innocent'.

A consequence is that if the general public is becoming hesitant to share their data, it may be detrimental for innovation. Transport policy and transport services rely on location-based data, and the absence such data limits the quality of service. Solutions to limit the spread of data exists, such as anonymization strategies or stricter data policies, however, they are not yet spread widely through society. Finding the right balance between distributing enough personal information while privacy and innovation is maintained will be the major issue in the future. Currently personal information too widely available, and the GDPR is the most serious attempt to restore the balance. However, the sweet spot has not yet been reached.

8 Big Data and the Centralisation of Market Power

It is increasingly being debated that companies that have achieved a dominant market position based on collecting large amounts of data should be under more supervision of the government[47] (Bourreau, De Streel, & Greaf, 2017; Sokol & Comerford, 2016; Stucke & Grunes, 2016). The massive size data-based companies like Amazon, Facebook and Google manage to obtain gives away the impression that there is a tendency for big-data based business models to lead to a concentration of market power in the hands of a few. This is based on the view that collecting large amount of data (personal and consumer behaviour data), for example by social networks, e-commerce platforms and search engines, gives these companies an unfair competitive advantage that can harm consumers and competition. While being big in itself is not an offense, it is the related consumer welfare loss that has brought these companies attention from regulators. This chapter examines the issue related to the market power resulting from big data.

While within the transportation sector there are no equivalents of Google and Facebook – data rich companies that amassed a massive market share in certain markets. Nevertheless, these companies could play an important role in the transport sector through the geographical component of the data they own. This section examines to what extent this general issue may be applicable to the transportation sector.

8.1 Does big data lead to reduced consumer welfare?

Big data has enabled the processing of much more (consumer) information than ever before. This has facilitated the emergence of a few big and well-known players in the data economy. These companies have gotten increasingly dominant by their sheer market size and through mergers and acquisitions, catching the eye of authorities.

The critique is that established data-rich companies sustain an unfair competitive advantage over others. This advantage leads to an overall loss of consumer welfare, for example when the dominant market position is used for slacking product quality, innovation or privacy protection, or for anti-competitive conduct. Sokol & Comerford (2016) discuss four ways in which big data may lead to an unfair competitive advantage and subsequently to a loss of consumer welfare: through (1) the extent to which operational efficiency is gained from big data (2) loss of quality and innovation, (3) loss of privacy and (4) offensive mergers to eliminate competition.

8.1.1 Operational efficiency: scale effects, networking effects and high entry barriers

Some note that the nature of big data allows for an operational efficiency based on scale effects and networking effects that makes it difficult for competitors to rival established companies. Scale effects refer to the benefits derived from the large scale of operation that large businesses

---

[47] See for example: https://www.theguardian.com/commentisfree/2018/apr/20/tech-monopoly-apple-facebook-data-extreme-content; https://www.theguardian.com/technology/2018/mar/14/tech-big-data-capitalism-give-wealth-back-to-people; https://www.nytimes.com/2017/01/08/technology/data-regulators-google-facebook-monopoly.html; http://nymag.com/selectall/2017/09/google-facebook-and-the-new-type-of-tech-monopoly.html;

can achieve. More users lead to more revenue, and more users leads to more data upon which the product can be improved. Networking effects refers to the phenomena that an additional user enhances the quality of the whole product or service. For example: social networking sites become more attractive the more users sign up; search engines are able to provide more accurate search results based on previous search queries; the more data from semi-automated vehicles is gathered by Tesla, the better are the self-driving vehicles made by Tesla. In addition, the barriers to enter data-based markets are considered to be relatively high. Using big data is a demanding exercise due to the availability and the costs involved with collecting and analysing big data. It thus takes time to effectively compete with existing companies.

Given that, time and time again, new innovative products are overthrowing companies with access to much bigger datasets suggests that scale and networking effects or high entry barriers are not sufficient to maintain a competitive edge. Slack, Snapchat, Instagram, Uber and Tinder – and originally Amazon, Facebook and Google as well – are examples of companies that found new ways to gain market power in data-driven business models.

Big data is everywhere, inexpensive and easy to collect (Sokol & Comerford, 2016, p. 1136). Many datasets are offered by the governments for free (Janssen, Charalabidis, & Zuiderwijk, 2012). New data is constantly created with every movement users make in space or in cyberspace. Without much effort purchasing patterns or mouse-clicks can be recorded. Analysis can be inexpensive and simple to use through ready-to-use and open source software available online. Moreover, having much extra data does not necessarily add much extra value. It is possible to relatively quickly identify what the most popular products are, which routes are most frequently travelled at what time or what services users prefer. Additionally, the lack of user data can be solved by purchasing information from third parties such as data brokers.

Moreover, big data is non-rivalrous and non-exclusive (Sokol & Comerford, 2016, p. 1137). A data point can be spread indefinitely and multiple companies can own the same data. When one company uses the data, it does not mean that another company cannot use the same data. Moreover, some data loses its worth over time quickly. User preferences and trends shift over time. With regards to travel information, most consumers wish to know what is happening right now rather than knowing what delays there were in the past.

8.1.2 Loss of quality and innovation

Big data enables companies to offer higher quality goods and services than their rivals, who have less access to data. The larger the quality gap, the less incentivized the established company is to innovate or to improve product quality. The reason for this is that reduced product quality does not lead to consumer loss for the leading company when the product is still far superior to the competitor's product. The revenue saved from 'not innovating' and thus 'not improving consumer well-being' could be for example used for marketing or extra profit. However, to what extent this happens in reality is highly unclear. The degree to which the loss of quality and innovation can be measured in relation the potential is difficult to establish. It is telling that no court of regulator has yet decided that market power has led to such a degradation in product quality and innovation (and hence consumer welfare loss) that the company should be fined. On top of that, companies need to keep innovating and improve product quality because of the volatility of the data economy. Companies that are not innovating are at danger of being blindsided by other companies offering new and better products.

8.1.3 Loss of privacy

In the big data economy, most products are offered for free because the consumers pay with their data instead of paying with currency. Privacy, in this regard, is a form of non-price competition. Companies compete by offering better privacy protection policies than other companies. However, some companies obtained such a market share that they are able to offer poor privacy policy without losing customers. This is especially problematic because people are generally not able to assess the value of their data, as described earlier in 7.1.1, 'The value of personal data'. (Sokol and Comerford 2017) note that this issue is the main issue when it comes to market power abuse.

8.1.4 Data Mergers

Mergers or company acquisition can become problematic when the result is that one company receives a dominant market competition. In particular, there have been mergers primarily motivated by gaining control of the dataset owned by another company. So far, antitrust agencies are well equipped to deal with assessing whether merges lead to unfair compaction. Recent statements by the European Commission regarding the Google/DoubleClick merger or the acquisition of WhatsApp by Facebook indicate that (1) sufficient choice for consumers remains to opt for a similar product from similar quality by a competing platform and (2) no data monopoly is gained as the type of data on which the companies are based remain accessible for competitors. However, the European Commission did not review any privacy related concerns.

8.2 Summary

While having big data certainly gives market power, having big data is not enough to maintain it. Despite 'the more big data the better', there remain plenty abilities and incentives for competitors with less data to challenge the market leaders. User preferences shift over time allowing for new products to gain popularity and there are various ways to develop new products without having to rely on scale and networking effects. The extent to which data power leads companies slacking and providing lesser quality products is difficult to establish, but seems unlikely (due to the volatility of the data market and the need to continuously innovate) and has never been the cause of conviction by antitrust authorities.

In addition, antitrust authorities are well equipped to track and fine companies for anti-competitive behaviour or mergers that lead to an unfair competitive advantage. And the same time, it is worth noting that the data market does bring new challenges to antithrust authorities. In particular, the assessment of market power based evaluating data availability and data value is difficult (Bourreau, de Streel, and Graef 2017).

Another issue is the fact that these data-rich companies often deal with personal data. Lacking privacy policies – which market power allows – does not lead to unfair competition and hence does not fall under the jurisdiction of antitrust agencies (Kennedy, 2016; Sokol & Comerford, 2016). This topic should be the concern of consumer protection authorities.

What impact do these developments have on the transport sector specifically? Compared to the tech sector, the data-rich players in the transport sector are not equally dominant. The legal classification of companies can make matters less clear. If judged as tech companies, the existing case law described earlier would not significantly hinder expansion of market power. If classified as a different sort of corporation, the situation is completely different.

For instance, ride service Uber has had its run-in with the European Court of Justice. This court ruled it to be primarily a transport company and not a provider of online services, subjecting it to different rules and regulations and making it more accountable to national law. This could impede its growth to ever become a truly dominant market player.

9 Conclusion

The purpose of this document is to identify the main economic and political issues involved with big data, and in particular those issues that are relevant for the usage of big data in the transportation sector.

A thing was identified as an issue when it poses a challenge or offers an opportunity for the usage of big data by economic and political actors.

9.1 Approach

The method used is desk-research. Primarily academic literature is reviewed to study how economic and political actors use big data and what challenges this poses. Where useful, reports or information from other reputable sources were used.

This document is structured as follows. The analysis framework is laid out in chapter 2. Economic issues are discussed in chapter 3 (firm level) and 4 (industry level), and political issues in chapter 5 (national level) and 6 (international level). The issue of privacy is examined in chapter 7, and chapter 8 deals with the question what market power is derived from big data.

9.2 Analysis framework

To grasp the complex domain of economics and politics, the first step was to identify the relevant actors. This was done in chapter 2.

The actors in the economic and political dimension are classified according to aspects of transport (i.e. demand, supply, externalities and governance) and according to roles in the data economy (i.e. data user, supplier or facilitator).

9.3 Findings: per chapter

Having identified the components of the economic and political dimension, the next step was to further elaborate on the economic issues by describing the relation between the transport stakeholders and the data actors (chapter 3). Transport data users and transport data suppliers use big data for improved situational awareness of the transport system, improving the capacity of transport networks, improving transportation services and facilitating the shift to sustainable transport. Concerns for achieving these benefits relate to big data quality, quality of the data system, and quality of the organisation.

The relations between the economic actors are examined in chapter 4, where the factors influencing the growth and innovation in the data economy are considered. The overall view is that European data economy lacks behind the American data economy, not only in terms of innovation, also regarding education. In general, US's technological innovation ecosystem is ahead of Europe. Moreover, big data usage in the economy causes negative externalities regarding the pressure on climate change. Currently, the IT ecosystem takes up more than 10% of the world's energy consumption. As the IT sector keeps growing, in particular with the advent of electric vehicles, we need to consider more efficient energy usage of data centres, IT manufacturing, end-use devices and IT infrastructure.

To grasp the political domain, in chapter 5, the application of big data by governments is considered. Governments utilize big data in improving organisational performance and in service provision and policy making. Subsequently, big data is applied in transport related government tasks, including transport planning, traffic monitoring and public transport provision.

To understand the power relations in the political domain, the arena where governments interact on the state level and compete and cooperate over the data economy was considered in chapter 6. This chapter revolves around the question what happens when data crosses the border and why governments may want to impact cross-border data flows. Governments want to control data flows to limit the negative consequences of data, e.g. preventing the misuse of personal or classified data. Another reason is to more easily carry out their task as supervisor, e.g. demanding local storage of tax or gambling data to make is easier to carry out control routines.

While governments seek to regulate data flows for the public good, imposing restrictions can have negative consequences for data-based innovation. Commentators note that current cross-border data legislation is too destructive for innovation, and the same aims can be achieved with less harsh restrictions on cross-border data flow (Graux, Ypma, and Foley 2017; Meltzer 2015; Ferracane 2017) – similar to how the WTO addresses international barriers to trade. However, an equivalent of the WTO for international data flows is not existing. The EU is currently addressing inefficiencies related to cross-border data flows. While the GDPR is well known, less known is that in June 2018, the EU reached a provisional political agreement on the free flow on non-personal data.

The most pressing issue for any actor handling data is examined in chapter 7: the concern about data and privacy. The main issues regarding the spread of personal data is that it is prone to being misused by individuals or groups that will process personal data unlawfully. Another concern relates to the negative consequences of the predictive capacity of big data. In some instances, the usage of big data for predictive analysis has led to exclusion, discrimination and autonomy loss of individuals. In transportation, many privacy issues has been observed as well, such as insufficient security levels, insufficient awareness of what constitutes personal information, and insufficient anonymization despite the anonymization technologies being there. There is still a lot of work to be done to establish the right framework with a proper trade-off between privacy protection and using personal information for business innovation.

The previous decades have seen the emergence of a few massive data rich companies, which raises the question of what power is inherent to owning big data. The market power of large companies becomes an issue when it causes harm to competition and consumers. Chapter 8 explores this issue. Despite ‘the more big data the better’, there remain plenty abilities and incentives for competitors with less data to challenge the market leaders. Given that many data-rich companies are US based, a better way to look at the cause is the technological innovation system rather than their market domination. So far, antitrust authorities are well-equipped to deal with market abuse by data rich companies, however, regulating the data economy does bring new challenges. It is quite difficult to determine market domination by determining data value and data accessibility by rivals. Another is issue that dominant players are able to use their market power to offer low privacy protection to consumers, leading to harm to consumers. Commentators suggest that this is not an antitrust issue and should be addresses by consumer protection authorities.

9.4 Findings: opportunities and challenges

The results from the chapters are reorganized into opportunities and challenges in Table 6. General issues are applicable to both private and public parties, after which specific issues for private and public parties are mentioned.

Table 6 Economic and political opportunities and challenges regarding the usage of big data identified in this report, divided into general and transport specific issues.

	Opportunities	**Concerns**
General Issues	**General** • More data improves predictive analysis and better products and services • Introduction of legislation facilitating cross-border flow of non-personal data • EU's policy 'road to digital market' promotes usage of big data technologies • Thanks to Massive Open Online Course, Big Data knowledge can be transferred to EU **Economy specific** • Innovation opportunities: market penetration, product development, market development and diversification • An initial capital to enter the market is low, because most of Big Data technologies are open sources and free to use. **Politics general** • Broadening evidence base for policy • Improvement of public service provision and • A lot of data available with no specific purpose	**General** • Privacy and the lack of regulation to deal privacy infringement • Lack of privacy policies offered by dominant market players • Issues inherent to the 5 V's of big data: bad data, 'dark data' i.e. collected data without a purpose • Costs involved with operating big data infrastructure • Organisational structure: effective usage of BD requires significant organisational efforts • General lack of skilled labour • Misconception that BDA is objective – it is inherently subjective and prone bad engineering and biases • Predictive analysis is prone to exclusion, discrimination and loss of autonomy • Digital exhaust, i.e. 'everything' is recorded • Privacy concerns makes consumers hesitant to share data **Economy specific** • The odds of the technical innovation system of BDA are in the favour of US compared to the EU • Limitations to the cross-border flow of non-personal data, however, EU policy is being discussed • Governments can impose various restrictions on cross-border data flows. **Politics specific**

		• Lack of interdepartmental integration, i.e. government 'siloing' • Prone to technocratic governance, i.e. improper usage BDA for too complicated topics • Private sector dependency
Transport related issues	**General** • Unexploited opportunity for anonymization technologies • Availability of transport related without a specific purpose **Economic specific** • Shared situational awareness of the transport network • Long term planning of the transport network • Real-time traffic management • Mitigating risks that degrade network capacity • Improving transport services through marketing activities, vehicle routing, service provision and scheduling • Supporting shift to sustainability **Politics specific** • Broadening evidence base for policy making and more thorough understanding of current and future transport • More safer and efficient transport systems • More and detailed data allows for more precise predictions regarding specific conditions/groups/events • Lots of transport related data available with no purpose yet • Allows for 'learning by doing' policy making • Allows to track things that could not be tracked before	**General** • Lack of public-private data sharing and the absence of a sufficient legal framework to enable this **Economic specific** • US has favourable market conditions for transport data innovations **Politics specific** • Planners deal with high expectations • Privacy issues for locational based data • Lack of data intergovernmental data sharing • BDA has been used for improving existing transport analysis, rather than developing new ways to analyse transport

References

ACEA. 2016a. "Access to Vehicle Data for Third-Party Services."

ACEA. 2016b. "ACEA Strategy Paper on Connectivity."

Aïvodji, U., Gambs, S., Huguet, M., & Killijian, M. (2016). Meeting points in ridesharing: A privacy-preserving. *Transportation research., 72*, 239-253. doi:10.1016/j.trc.2016.09.017

Aïvodji, Ulrich Matchi, Sébastien Gambs, Marie José Huguet, and Marc Olivier Killijian. 2016. "Meeting Points in Ridesharing: A Privacy-Preserving Approach." *Transportation Research Part C: Emerging Technologies* 72: 239–53. https://doi.org/10.1016/j.trc.2016.09.017.

Akter, Shahriar, and Samuel Fosso Wamba. 2016. "Big Data Analytics in E-Commerce: A Systematic Review and Agenda for Future Research." *Electronic Markets* 26 (2): 173–94. https://doi.org/10.1007/s12525-016-0219-0.

Anda, Cuauhtemoc, Alexander Erath, and Pieter Jacobus Fourie. 2017. "Transport Modelling in the Age of Big Data." https://doi.org/10.1080/12265934.2017.1281150org/10.1080/12265934.2017.1281150.

Andrew Zolli. 2018. "After Big Data: The Coming Age of 'Big Indicators.'" *Stanford Social Innovation Review*, 2018. https://ssir.org/articles/entry/after_big_data_the_coming_age_of_big_indicators.

Avoine, Gildas, Luca Calderoni, Jonathan Delvaux, Dario Maio, and Paolo Palmieri. 2014. "Passengers Information in Public Transport and Privacy: Can Anonymous Tickets Prevent Tracking?" *International Journal of Information Management* 34 (5). Elsevier Ltd: 682–88. https://doi.org/10.1016/j.ijinfomgt.2014.05.004.

Batty, Michael. 2013. "Big Data, Smart Cities and City Planning." *Dialogues in Human Geography* 3 (3): 274–79. https://doi.org/10.1177/2043820613513390.

Beiraõ, Gabriela, and J A Sarsfield Cabral. 2007. "Understanding Attitudes towards Public Transport and Private Car: A Qualitative Study." *Transport Policy* 14: 478–89. https://doi.org/10.1016/j.tranpol.2007.04.009.

Belcastro, Loris, Fabrizio Marozzo, Domenico Talia, and Paolo Trunfio. 2016. "Using Scalable Data Mining for Predicting Flight Delays." *ACM Transactions on Intelligent Systems and Technology* 8 (1): 1–20. https://doi.org/10.1145/2888402.

Bertot, John Carlo, Ursula Gorham, Paul T. Jaeger, Lindsay C. Sarin, and Heeyoon Choi. 2014. "Big Data, Open Government and e-Government: Issues, Policies and Recommendations." *Information Polity* 19: 5–16. https://doi.org/10.3233/IP-140328.

Bouma, K. (2016, June 15). *Niet alles wat mogelijk is moet je willen*. Retrieved from De Groene Amsterdammer: https://www.groene.nl/artikel/niet-alles-wat-mogelijk-is-moet-je-willen

Bourreau, M., De Streel, A., & Greaf, I. (2017). *Big data and competition policy: market power, personalised pricing and advertising.* Brussel: Centre on Regulation in Europe (CERRE).

Bourreau, Marc, Alexandre de Streel, and Inge Graef. 2017. "Big Data and Competition Policy: Market Power, Personalised Pricing and Advertising." *Centre on Regulation in Europe (CERRE)*. Vol. 32. https://doi.org/10.2139/ssrn.2920301.

Bryson, John;, Barbara Crosby, and Laura Bloomberg. 2014. "Public Value Governance: Moving Beyond Traditional Public Administration and the New Public Management." *Public Administration Review* 74 (4): 445–56. https://doi.org/10.1111/puar.12238.

Bucherer, Eva, Uli Eisert, and Oliver Gassmann. 2012. "Towards Systematic Business Model Innovation: Lessons from Product Innovation Management." *Creativity and Innovation Management* 21 (2): 183–98. https://doi.org/10.1111/j.1467-8691.2012.00637.x.

Cachon, Gérard P, Kaitlin M Daniels, and Ruben Lobel. 2017. "The Role of Surge Pricing on a Service Platform with Self-Scheduling Capacity." *Manufacturing & Service Operations Management* 19 (3): 368–84. https://doi.org/10.1287/msom.2017.0618.

Callon, Michel. 1998. "Introduction: The Embeddedness of Economic Markets in Economics." *The Sociological Review* 46 (1_suppl): 1–57. https://doi.org/10.1111/j.1467-954X.1998.tb03468.x.

Cattaneo, G, G Micheletti, D Osimo, and K Jakimowicz. 2018. "The European Data Market Monitoring Tool Report."

Cottrill, Caitlin D, and Sybil Derrible. 2014. "Leveraging Big Data for the Development of Transport Sustainability Indicators." *Journal of Urban Technology* 0 (1): 45–64. https://doi.org/10.1080/10630732.2014.942094org/10.1080/10630732.2014.942094.

Crawford, Kate, and Jason Schultz. 2014. "Big Data and Due Process - Toward a Framework To Redress Predictive Privacy Harms." *Boston College Law Review* 55 (1): 93–128. https://doi.org/10.1525/sp.2007.54.1.23.

Cruickshanks, S., & Waterson, B. (2012). Will privacy concerns associated with future transport systems restrict the public's freedom of movement? *Procedia - Social and Behavioral Sciences, 48*, 941-950. doi:doi.org/10.1016/j.sbspro.2012.06.1071

Cruickshanks, Scott, and Ben Waterson. 2012. "Will Privacy Concerns Associated with Future Transport Systems Restrict the Public's Freedom of Movement?" *Procedia - Social and Behavioral Sciences* 48. Elsevier B.V.: 941–50. https://doi.org/10.1016/j.sbspro.2012.06.1071.

Deibert, Ron. 2015. "The Geopolitics of Cyberspace after Snowden." *Current History* 114 (768): 9–15.

EDISON. (2016). *Existing educational and training resources.* EDISON.

Erevelles, Sunil, Nobuyuki Fukawa, and Linda Swayne. 2016. "Big Data Consumer Analytics and the Transformation of Marketing." *Journal of Business Research* 69: 897–904. https://doi.org/10.1016/j.jbusres.2015.07.001.

Eriksson, Louise, and Torkel Bjørnskau. 2012. "Acceptability of Traffic Safety Measures with Personal Privacy Implications." *Transportation Research Part F: Traffic Psychology and Behaviour* 15 (3). Elsevier Ltd: 333–47. https://doi.org/10.1016/j.trf.2012.02.006.

Eubanks, V. (2018). *Automating inequality.* St. Martin's Press.

European Commission. 2017. "Free Flow of Non-Personal Data" 2014: 2.

everis Benelux. 2018. *Study on Data Sharing between Companies in Europe*. European Union. https://doi.org/10.2759/354943.

Ferracane, Martina F. 2017. "Restrictions on Cross-Border Data Flows: A Taxonomy." *ECIPE Working Paper*, no. 1: 1–27.

Fries, R., Gahrooei, M., Chowdhury, M., & Conway, A. (2012). Meeting privacy challenges while advancing intelligent transportation systems. *Transportation Research Part C: Emerging Technologies, 25*, 34-45. doi:10.1016/j.trc.2012.04.002

Fu, Qian, and John M Easton. 2017. "Understanding Data Quality: Ensuring Data Quality by Design in the Rail Industry." In *2017 IEEE International Conference on Big Data (Big Data)*, 1117:3792–99. IEEE. https://doi.org/10.1109/BigData.2017.8258380.

Furth, Peter G, Brendon J Hemily, Theo H J Muller, and James G Strathman. 2003. "Uses of Archived AVL-APC Data to Improve Transit Performance and Management: Review and Potential - TCRP Web Document 23 (Project H-28)." http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_webdoc_23.pdf.

Gans, Gerhard De Koning, and JH Hoepman. 2008. "A Practical Attack on the MIFARE Classic." *Smart Card Research and Advanced Applications* 5189 (September): 267–82. https://doi.org/http://dx.doi.org/10.1007/978-3-540-85893-5_20.

Ghasemzadeh, Moein, Benjamin C M Fung, Rui Chen, and Anjali Awasthi. 2014. "Anonymizing Trajectory Data for Passenger Flow Analysis." *Transportation Research Part C* 39: 63–79. https://doi.org/10.1016/j.trc.2013.12.003.

Ghofrani, Faeze, Qing He, Rob M.P. Goverde, and Xiang Liu. 2018. "Recent Applications of Big Data Analytics in Railway Transportation Systems: A Survey." *Transportation Research Part C: Emerging Technologies* 90 (March). Elsevier: 226–46. https://doi.org/10.1016/j.trc.2018.03.010.

Giest, S. (2017). Big data for policymaking: fad or fasttrack? *Policy Science, 50*, 367-382. doi:10.1007/s11077-017-9293-1

Giest, Sarah. 2017. "Big Data for Policymaking: Fad or Fasttrack?" *Policy Sciences* 50 (3). Springer US: 367–82. https://doi.org/10.1007/s11077-017-9293-1.

Glaeser, Edward L., Scott Duke Kominers, Michael Luca, and Nikhil Naik. 2018. "Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life." *Economic Inquiry* 56 (1): 114–37. https://doi.org/10.1111/ecin.12364.

Godel, M, A Joshi, R Bond, J Cabello, and D Fernandez. 2016. "Report: Facilitating Cross Border Data Flow in the Digital Single Market." https://doi.org/10.2759/342893.

Gohar, Moneeb, Muhammad Muzammal, and Arif Ur Rahman. 2018. "SMART TSS: Defining Transportation System Behavior Using Big Data Analytics in Smart Cities." *Sustainable Cities and Society* 41 (May). Elsevier: 114–19. https://doi.org/10.1016/j.scs.2018.05.008.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography, 3*(3), 255-261. doi:10.1177/2043820613513121

Graham, Mark, and Taylor Shelton. 2013. "Geography and the Future of Big Data, Big Data and the Future of Geography." *Dialogues in Human Geography* 3 (3): 255–61. https://doi.org/10.1177/2043820613513121.

Graux, Hans, Patricia Ypma, and Paul Foley. 2017. "Cross-Border Data Flow in the Digital Single Market : Study on Data Location Restrictions." https://doi.org/10.2759/92602.

Groeve, Tom De, Karmen Poljansek, Daniele Ehrlich, and Christina Corbane. 2014. *Current Status and Best Practices for Disaster Loss Data Recording in EU Member States*. https://doi.org/10.2788/18330.

Halvorsen, Anne (Anne Fire). 2015. "Improving Transit Demand Management with Smart Card Data : General Framework and Applications." Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/99543.

Hee, K., Mushtaq, N., özmen, H., Rosselli, M., Zicari, R., Hong, M., . . . Teoh, T. (2018). *Deliverable 1.1: Understanding and Mapping big data in the Transport Sector.* Retrieved from Leveraging Big Data to Manage Transport Operations (LeMO): https://lemo-h2020.eu/newsroom/2018/5/13/deliverable-11-understanding-big-data-in-transport-sector

Hekkert, M, S Negro, G Heimriks, and R Harmsen. 2011. "Technological Innovation System Analysis: A Manual for Analysts." *Utrecht University - Faculty of Geosciences*.

Hekkert, M.P., R.A.A. Suurs, S.O. Negro, S. Kuhlmann, and R.E.H.M. Smits. 2007. "Functions of Innovation Systems: A New Approach for Analysing Technological Change." *Technological Forecasting and Social Change* 74 (4): 413–32. https://doi.org/10.1016/j.techfore.2006.03.002.

Hibbs, John. 2003. *Transport Economics and Policy: A Practical Analysis of Performance, Efficiency and Marketing Objectives*. London: Kogan Page. https://s3.amazonaws.com/academia.edu.documents/39313531/Transport_Economics_and_Policy.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1531135375&Signature=gPZFmrwLhnE9GVA7Mds7jglQnp4%3D&response-content-disposition=inline%3B filename%3DTransport_Economi.

Höchtl, J., Parycek, P., & Schölhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Compputing and Electronic Commerce, 26*(1-2), 147-169. doi:10.1080/10919392.2015.1125187

Höchtl, Johann, Peter Parycek, and Ralph Schöllhammer. 2016. "Big Data in the Policy Cycle: Policy Decision Making in the Digital Era." *Journal of Organizational Computing and Electronic Commerce* 26 (1–2). Taylor & Francis: 147–69. https://doi.org/10.1080/10919392.2015.1125187.

Hommes, E, and M Holmner. 2013. "Intelligent Transport Systems : Privacy , Security and Societal Considerations within the Gauteng Case Study E Hommes and M Holmner." *Information Ethics*, no. 46: 192–206.

Hu, Shengneng, and Wei Liu. 2017. "Research on the Application of Big Data in Intelligent Transportation System." *Revista de La Facultad de Ingeniería U.C.V* 32 (5): 517–24. https://doi.org/10.1109/ICCNEA.2017.35.

Hubauer, Thomas, Steffen Lamparter, Mikhail Roshchin, Nina Solomakhina, and Stuart Watson. 2013. "Analysis of Data Quality Issues in Real-World Industrial Data." In *Annual Conference of the Prognostics and Health Management Society*. New Orleans. http://ftp.phmsociety.org/sites/phmsociety.org/files/phm_submission/2013/phmc_13_067.pdf.

IBM. 2014. "Big Data and Analytics in Travel and Transportation." Somers, New York. http://www-07.ibm.com/au/pdf/Travel_Transportation.pdf.

IDC, and Open Evidence. 2017. "The European Data Market Study: Final Report."

Issenberg, S. (2012). *The victory lab: the secret science of winning campaigns.* New York: NY: Crown.

Jamshidi, Ali, Shahrzad Faghih-Roohi, Siamak Hajizadeh, Alfredo Núñez, Robert Babuska, Rolf Dollevoet, Zili Li, and Bart De Schutter. 2017. "A Big Data Analysis Approach for Rail Failure Risk Assessment." *Risk Analysis* 37 (8): 1495–1507. https://doi.org/10.1111/risa.12836.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management, 29*(4), 258-268. doi:doi.org/10.1080/10580530.2012.716740

Keller, Richard M, Shubha Ranjan, Mei Y Wei, and Michelle M Eshow. 2016. "Semantic Representation and Scale-up of Integrated Air Traffic Management Data." In *Proceedings of the International Workshop on Semantic Big Data - SBD '16*, 1–6. New York, New York, USA: ACM Press. https://doi.org/10.1145/2928294.2928296.

Kennedy, J. (2016, March). The myth of data monopoly: why antitrust concerns about data are overblown. *Information Technology & Innovation Foundation*.

Kerr, I., & Earle, J. (2013). Prediction, preemption, presumption: how big data threathens big picture privacy. *Stanford Law Review, 66*(65). Retrieved 7 12, 2018, from https://www.stanfordlawreview.org/online/privacy-and-big-data-prediction-preemption-presumption/

Khan, Sakib M., Linh B. Ngo, Eric A. Morris, Kakan Dey, and Yan Zhou. 2017. *Social Media Data in Transportation*. *Data Analytics for Intelligent Transportation Systems*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-809715-1.00011-0.

Khan, Sakib M., Mizanur Rahman, Amy Apon, and Mashrur Chowdhury. 2017. *Characteristics of Intelligent Transportation Systems and Its Relationship With Data Analytics*. *Data Analytics for Intelligent Transportation Systems*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-809715-1.00001-8.

Kim, G., Trimi, S., & Chung, J. (2014). Big-data applications in the government sector. *Communications of the ACM, 57*(3), 78-85. doi:10.1145/2500873

Kim, Gang-Hoon, Silvana Trimi, and Ji-Hyong Chung. 2014. "Big-Data Applications in the Government Sector." *Communications of the ACM* 57 (3): 78–85. https://doi.org/10.1145/2500873.

Kim, H., & Cho, J. (2017). Data governance framework for big data implementation with a case of Korea. *IEEE 6th International Congress on Big Data*, (pp. 384-391).

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 1-14.

Kitchin, Rob. 2014. "The Real-Time City? Big Data and Smart Urbanism." *GeoJournal* 79 (1): 1–14. https://doi.org/10.1007/s10708-013-9516-8.

Kurbalija, J. (2017, May 9). *The impact of (big) data on geopolitics, negotiations, and the diplomatic modus operandi*. Retrieved 7 26, 2018, from Diplo: https://www.diplomacy.edu/blog/impact-big-data-geopolitics-negotiations-and-diplomatic-modus-operandi

Lathia, Neal, Chris Smith, Jon Froehlich, and Licia Capra. 2013. "Individuals among Commuters: Building Personalised Transport Information Services from Fare Collection Systems." *Pervasive and Mobile Computing* 9 (5): 643–64. https://doi.org/10.1016/j.pmcj.2012.10.007.

Laurent, A. (2004). *Understanding open source and free software licensing: guide to navigating licensing issues in existing & new software.* O'Reilly Media, Inc.

Lederman, J., Taylor, B., & Garrett, M. (2016). A private matter: the implications of privacy regulations for intelligent transportation systems. *Transportation Planning and Technology*, 1-21. doi:10.1080/03081060.2015.1127537

Lederman, Jaimee, Brian D. Taylor, and Mark Garrett. 2016. "A Private Matter: The Implications of Privacy Regulations for Intelligent Transportation Systems." *Transportation Planning and Technology* 39 (2): 115–35. https://doi.org/10.1080/03081060.2015.1127537.

Li, Hongfei, Dhaivat Parikh, Qing He, Buyue Qian, Zhiguo Li, Dongping Fang, and Arun Hampapur. 2014. "Improving Rail Network Velocity: A Machine Learning Approach to Predictive Maintenance." *Transportation Research Part C* 45: 17–26. https://doi.org/10.1016/j.trc.2014.04.013.

Lillo, R. 2018. "Data and Analytics Framework. How Public Sector Can Profit from Its Immense Asset, Data." In *Traffic Mining Applied to Police Activities*, edited by Fabio Leuzzi and Stefano Ferilli, 728:3–10. https://doi.org/10.1007/978-3-319-75608-0.

Lind, Mikael, Mikael Hägg, Ulf Siwe, and Sandra Haraldson. 2016. "Sea Traffic Management – Beneficial for All Maritime Stakeholders." In *6th Transport Research Arena April 18-21, 2016*, 14:183–92. https://doi.org/10.1016/j.trpro.2016.05.054.

Liu, Weng Kun, and Chia Chun Yen. 2016. "Optimizing Bus Passenger Complaint Service through Big Data Analysis: Systematized Analysis for Improved Public Sector Management." *Sustainability* 8 (12): 1–21. https://doi.org/10.3390/su8121319.

Ma, Xiaolei, Yao-Jan Wu, Yinhai Wang, Feng Chen, and Jianfeng Liu. 2013. "Mining Smart Card Data for Transit Ridersâ€™ Travel Patterns." https://doi.org/10.1016/j.trc.2013.07.010.

Malgieri, Gianclaudio, and Bart Custers. 2018. "Pricing Privacy." *Computer Law and Security Review* 34 (2): 1–22. https://doi.org/doi.org/10.1016/j.clsr.2017.08.006.

Margretts, M., & Dunleavy, P. (2013). The second wave of digital-era governance: a quasi-paradigm for government on the Web. *Philosophical Transactions of the Royal Society A, 371*, 1-17.

Martin, P, Y Feng, and X Wang. 2003. "Detector Technology Evaluation."

Meltzer, Joshua Paul. 2015. "The Internet, Cross-Border Data Flows and International Trade." *Asia and the Pacific Policy Studies* 2 (1): 90–102. https://doi.org/10.1002/app5.60.

Mergel, I., Rethemeyer, K., & Isett, K. (2016). Big Data in Public Affairs. *Public Administration review, 76*(6), 928-937. doi:10.1111/puar.12625

Mergel, Ines, R. Karl Rethemeyer, and Kimberley Isett. 2016. "Big Data in Public Affairs." *Public Administration Review* 76 (6): 928–37. https://doi.org/10.1111/puar.12625.

Miles, Samantha. 2017. "Stakeholder Theory Classification: A Theoretical and Empirical Evaluation of Definitions." *Journal of Business Ethics* 142 (3): 437–59. https://doi.org/10.1007/s10551-015-2741-y.

Mills, M. 2013. "The Cloud Begins with Coal: Big Data, Big Networks, Big Infrastructure and Big Power." https://doi.org/10.1227/01.NEU.0000137782.18834.E1.

Milne, Dave, and David Watling. 2018. "Big Data and Understanding Change in the Context of Planning Transport Systems." *Journal of Transport Geography*, no. November. Elsevier: 1–10. https://doi.org/10.1016/j.jtrangeo.2017.11.004.

Miltner, O. (2018, May 2). *Deploying big data to defend the US*. Retrieved August 2, 2018, from Ozy: https://www.ozy.com/acumen/deploying-big-data-to-defend-the-us/82416

Milutinovic, Milica, Koen Decroix, Vincent Naessens, and Bart Decker. 2015. "Privacy-Preserving Public Transport Ticketing System." In *29th IFIP Annual Conference on Data and Applications Security and Privacy (DBSEC), Jul 2015, Fairfax, VA, United States*, 1–17.

Minghetti, Valeria, and Dimitrios Buhalis. 2010. "Digital Divide in Tourism." *Journal of Travel Research* 49 (3): 267–81. https://doi.org/10.1177/0047287509346843.

Montjoye, Y., Hidalgo, C., Verleysen, M., & Blondel, V. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports, 3*(1376), 1-5. doi:DOI: 10.1038/srep01376

Natvig, Marit K. 2009. "SMARTFREIGHT–A Generic Approach to ITS for Smart Urban Freight Distribution." *16th ITS World Congress and Exhibition on Intelligent Transport Systems and ServicesITS AmericaERTICOITS Japan*, no. February.

Nik Rouda. 2014. "Getting Real About Big Data: Build Versus Buy." The Enterprise Strategy Group. http://www.oracle.com/us/corporate/analystreports/esg-getting-real-bigdata-2228170.pdf.

Njuguna, Christopher, and Patrick Mcsharry. 2017. "Constructing Spatiotemporal Poverty Indices from Big Data ☆." *Journal of Business Research* 70: 318–27. https://doi.org/10.1016/j.jbusres.2016.08.005.

Nofi, Albert. 2000. “Defining and Measuring Shared Situational Awareness.” Alexandria, Virginia. https://www.cna.org/cna_files/pdf/D0002895.A1.pdf.

Nouskalis, G. 2011. “Biometrics, e-Identity, and the Balance between Security and Privacy: Case Study of the Passenger Name Record (PNR) System.” *TheScientificWorldJournal* 11: 474–77. https://doi.org/10.1100/tsw.2011.48.

O’Neil, C. (2016). *Weapons of math destruction.* Broadway Books.

Ortúzar, Juan de Dios, and Luis G. Willumsen. 2011. *Modelling Transport*. *Modelling Transport*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119993308.

Papacharalampous, A. E., Stefan Hovelynck, O. Cats, J. W. Lankhaar, W. Daamen, N. Van Oort, and J. W.C. Van Lint. 2015. “Multi-Modal Data Fusion for Big Events [Research News].” *IEEE Intelligent Transportation Systems Magazine* 7 (4): 5–10. https://doi.org/10.1109/MITS.2015.2474940.

Parcu, Pier Luigi, Timothy Brennan, and Victor Glass. 2018. *The Contribution of the Postal and Delivery Sector: Between E-Commerce and E-Substitution*. Edited by Pier Luigi Parcu, Timothy Brennan, and Victor Glass. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-70672-6.

Potoglou, Dimitris, Neil Robinson, Chong W. Kim, Peter Burge, and Richard Warnes. 2010. “Quantifying Individuals’ Trade-Offs between Privacy, Liberty and Security: The Case of Rail Travel in UK.” *Transportation Research Part A: Policy and Practice* 44 (3). Elsevier Ltd: 169–81. https://doi.org/10.1016/j.tra.2009.12.006.

Rubinstein, I. (2013). Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law*, 1-14.

Rubinstein, I. S. 2013. “Big Data: The End of Privacy or a New Beginning?” *International Data Privacy Law* 3 (2): 74–87. https://doi.org/10.1093/idpl/ips036.

Sadeghi, A, I Visconti, and C Wachsmann. 2008. “User Privacy in Transport Systems Based on RFID E-Ticket.” Edited by C Bettini, A Jajodi, P. Samarati, and X.S. Wang. *PiLBA, Volume 397 of CEUR Workshop Proceedings. CEUR-WS.Org*.

Sadilek, Adam, and John Krumm. n.d. “Far Out: Predicting Long-Term Human Mobility.” Accessed July 13, 2018. www.aaai.org.

Sanders, N. (2016). How to use big data to drive your supply chain. *California Management Review, 58*(3), 26-48.

Santos, Georgina, Hannah Behrendt, Laura Maconi, Tara Shirvani, and Alexander Teytelboym. 2010. “Part I: Externalities and Economic Policies in Road Transport.” *Research in Transportation Economics* 28 (1). Elsevier Ltd: 2–45. https://doi.org/10.1016/j.retrec.2009.11.002.

Schroeder, Ralph. 2016. “Big Data Business Models: Challenges and Opportunities.” *Cogent Social Sciences* 2. https://doi.org/10.1080/23311886.2016.1166924.

Shi, Qi, and Mohamed Abdel-Aty. 2015. “Big Data Applications in Real-Time Traffic Operation and Safety Monitoring and Improvement on Urban Expressways.” *Transportation Research Part C: Emerging Technologies* 58. Elsevier Ltd: 380–94. https://doi.org/10.1016/j.trc.2015.02.022.

Siemens. 2012. “Toll, Traffic Jam or Shuttle Bus? Now Drivers Using a Highway Outside Tel Aviv Can Decide for Themselves.” *Como*, September 2012. https://www.mobility.siemens.com/mobility/global/SiteCollectionDocuments/en/webfeature-green-mobility/files/article/Fast_Lane_article_como09_2012_EN.pdf.

Singh, D, C Vishnu, and C K Mohan. 2017. “Visual Big Data Analytics for Traffic Monitoring in Smart City.” *15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, 886–91. https://doi.org/10.1109/ICMLA.2016.177.

Sokol, D., & Comerford, R. (2016). Antitrust and regulating Big Data. *Geo. Mason. L. REv., 23*(5), 1129-1160.

Sokol, D.Daniel, and Roisin Comerford. 2017. “Antitrust and Regulating Big Data.” *George Mason Law Review* 23: 1–33. https://doi.org/10.3366/ajicl.2011.0005.

Solakivi, Tomi, and Lauri Ojala. 2017. “Determinants of Carrier Selection: Updating the Survey Methodology into the 21 St Century.” *Transportation Research Procedia* 25 (00): 511–30. https://doi.org/10.1016/j.trpro.2017.05.433.

Steenbruggen, John, Peter Nijkamp, Jan M Smits, and Michel Grothe. 2011. “Traffic Incident Management: A Common Operational Picture to Support Situational Awareness of Sustainable Mobility.” Amsterdam.

Stucke, M., & Grunes, A. (2016). *Big data and competition policy.* Oxford: Oxford University Press.

Sunstein, C. (2015). The Ethics of Nudging. *Yale Journal on Regulation, 32*(2), 413-450.

TechAmerica. (2015). *Demystifying big data - a practical guide to transforming the business of government.* Washington, DC: TechAmerica Foundation.

The Economist. 2018. “What If People Were Paid for Their Data.” 2018.

UNICA. 2017. “Benutting Restwarmte Uit Datacenters.”

Wu, Pei-Ju, and Yi-Ching Chen. 2017. “Big Data Analytics for Transport Systems to Achieve Environmental Sustainability.” In *The 2017 IEEE International Conference on Applied System Innovation IEEE-ICASI 2017*, 264–67. https://doi.org/10.1109/ICASI.2017.7988401.

Yao, M., Rice, R., & Wallis, K. (2007). Predicting User Concerns About Online Privacy. *Journal of the American Society for Information Science and Technology, 58*(5), 710-722. doi:10.1002/asi.20530

Zaidi, Kamran, and Muttukrishnan Rajarajan. 2015. “Vehicular Internet: Security & Privacy Challenges and Opportunities.” *Future Internet* 7 (3): 257–75. https://doi.org/10.3390/fi7030257.

Zaman, Ibna, Kayvan Pazouki, Rose Norman, Shervin Younessi, and Shirley Coleman. 2017. “Challenges and Opportunities of Big Data Analytics for Upcoming Regulations and Future

Transformation of the Shipping Industry." *Procedia Engineering* 194: 537–44. https://doi.org/10.1016/j.proeng.2017.08.182.

Zeng, Gang. 2015. "Application of Big Data in Intelligent Traffic System." *IOSR Journal of Computer Engineering* 17 (1): 2278–2661. https://doi.org/10.9790/0661-17160104.

Zhang, Junping, Fei-Yue Wang, Kunfeng Wang, Wei-Hua Lin, Xin Xu, and Cheng Chen. 2011. "Data-Driven Intelligent Transportation Systems: A Survey." *IEEE Transactions on Intelligent Transportation Systems* 12 (4): 1624–39. https://doi.org/10.1109/TITS.2011.2158001.

Zhong, Ray Y, Stephen T Newman, George Q Huang, and Shulin Lan. 2016. "Big Data for Supply Chain Management in the Service and Manufacturing Sectors: Challenges, Opportunities, and Future Perspectives." *Computers & Industrial Engineering* 101 (November): 572–91. https://doi.org/10.1016/j.cie.2016.07.013.