LEVERAGING BIG DATA FOR MANAGING TRANSPORT OPERATIONS

Deliverable 4.2 **Horizontal Analysis and Socio-Economic Impacts**

Minsung Hong[1], Rajendra Akerkar[1], Tharsis Teoh[2], Timo Eichhorn[3], Julien Debussche[4], Jayant Sangwan[5]

Western Norway Research Institute[1],

Panteia B.V.[2],

Goethe-University Frankfurt[3]

Bird&Bird[4]

CORTE[5]

February 2020

Work Package 4

Project Coordinator

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

Horizon 2020 Research and Innovation Programme

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

Distribution level	Public (P)
Due date	15/03/2020 (Planned due date: 28/02/2020 (M28))
Sent to coordinator	13/03/2020
No. of document	D4.2
Title	*Horizontal analysis and socio-economic impacts*
Status & Version	*0.5*
Work Package	*4: Horizontal analysis and road-mapping*
Related Deliverables	*D4.1, D4.3 and D4.4*
Leading Partner	*WNRI*
Leading Authors	*Minsung Hong, WNRI* *Rajendra Akerkar, WNRI* *Tharsis Teoh, Panteia* *Timo Eichhorn, GUF* *Julien Debussche, B&B* *Jayant Sangwan, CORTE*
Contributors	*Anna Maria Urbaniak-Brekke, WNRI*
Reviewers	*Todor Ivanov, GUF*
Keywords	*Horizontal analysis, Big Data, opportunities, case study, technological, policy, legal, ethical, social, economic, political and environmental issues, socio-economic impact analysis, comparative analysis, natural language processing, visualization*

Executive summary

This deliverable presents a horizontal analysis of the case studies to produce constructive findings and suggestions on the prerequisites of successfully implementing big data in the transport sector.

This deliverable particularly looks at “opportunities/interventions”, which were defined as ways to promote the benefits of big data and foster the use of big data in transport sector (See D4.1) and introduces some of the interventions that could help overcome the barriers and limitations in D.4.1. The opportunities and interventions are co-related to the issues of the barriers and limitations that were identified in the case studies under the project and to the aspects examined in work packages 1, 2, and 3, as well as in Task 4.1.

The findings and suggestions in the deliverable are examined from a socio-economic point of view by not only a comparative analysis on the tables in D4.1 and this deliverable but also NLP (Natural Language Processing) and visualisation techniques. Some visualisation results will be uploaded in the LeMO website to provide interactive interfaces giving people more deep information about the analysis.

The used dataset includes total 243 items of around 62 issues as barriers/limitations and opportunities/interventions on the use of big data in the transport sector. The findings will be used in Tasks 4.3 and 4.4 to provide research and policy roadmaps.

Table of contents

List of Figures

List of Tables

Glossary

Abbreviation	Expression
AIS	Automatic Identification System
API	Application Programming Interface
BDA	Big Data Analytic
BDV	Big Data Value
BDVA	Big Data Value Association
CCTV	Closed-circuit television
CS	Case Study
EDISON	Education for Data Intensive Science to Open New science frontiers
EDSA	European Data Science Academy
ENISA	European Union Agency for Network and Information Security
GDPR	General Data Protection Regulation
GPS	Global Positioning System
HPDA	High Performance Data Analytics
ICT	Information and Communications Technology
IDP	Industrial Data Platform
IoT	Internet of Things
ISO	International Organization for Standardization
ITS	Intelligent transportation system
KPI	Key Performance Indicator
LeMO	Leveraging Big Data for Managing Transport Operations

ML	Machine Learning
MOD	Mobility On Demand
NIS	Network and Information Security
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
NOESIS	Novel Decision Support tool for Evaluating Strategic Big Data investments in Transport and Intelligent Mobility Services
PaaSword	A Holistic Data Privacy And Security By Design Platform-As-A Service Framework
PAPAYA	PlAtform for PrivAcY preserving data Analytics
PDP	Personal Data Platform
PRISMACLO	PRIvacy & Security MAintaining Services in the CLOUD
PSI	Public Sector Information
RDP	Research Data Platform
ROI	Return On Investment
SFIA	Skills Framework for the Information Age
SIA	Socio-economic Impact Analysis
SME	Small and Medium-sized Enterprise
TT	Transforming Transport
UDP	Urban/City Data Platform

1 Introduction

1.1 Abstract

In this deliverable, we evaluate "opportunities" for utilisation of big data in transport operations and introduce some interventions that can help overcoming some of the barriers and limitations discussed in D4.1. Previously, the LeMO project performed seven case studies to investigate issues related to technological, legal, ethical, social, environmental, economic, policy and political aspects in the use of big data in transport. Thereafter, the barriers and limitations identified from the case studies and other initiatives (i.e., Evaluating Strategic Big Data investments in Transport and Intelligent Mobility Services (NOESIS)[1] and Transforming Transport (TT)[2]) were discussed in Task 4.1.

This deliverable will focus on the opportunities related to the use of big data in transport, which will be linked to and considered in the context of technological, legal, ethical, social, environmental, economic, policy and political aspects. In addition, a socio-economic impact analysis will be introduced to evaluate the barriers and opportunities, and some interventions will be introduced to overcome the barriers and limitations. These results will help in devising a set of best practices that will be covered in deliverables D4.3 and D4.4.

1.2 Purpose of the document

The purpose of this deliverable is to undertake a horizontal analysis of the case studies to produce constructive findings and suggestions on the prerequisites of successful use of big data implementation in the transport sector.

To do this, the deliverable will identify co-relations between investigated barriers/limitations & opportunities/interventions and assess the impact of its findings by analysing them from a socio-economic point of view. This deliverable will feed into other deliverables on the development of research and policy roadmaps, i.e. D4.3 and D4.4.

1.3 Target audience

The target audience for this deliverable is:

- Partners and Advisory & Reference Group in the LeMO project;
- European Commission;
- EU Parliament;

[1] NOESIS: https://noesis-project.eu/
[2] Transforming Transport: https://transformingtransport.eu/

- Horizon 2020 projects and related transport projects (cfr. clustering activities);
- Organisations and experts involved in the LeMO case studies;
- Public and private transport organisations;
- Authorities (regional and national level) that develop and enforce policies and legislations.

2 Methodology

This section recalls the definition of terms used in this work package as mentioned in D4.1 [19] and briefly introduces the practical methodologies, which guide the investigation of the technical issue and socio-economic impact analysis, in LeMO project.

Definition of terms in WP4

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

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

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

Notation expressing the issues at the rest of this document: Notation structures of reference are different according to the format “OPT (Opportunities)-Aspect-Major issue (Initial)-Sub issue (Number)”.

2.1 Big data reference model for technical and technological issue

As value chains have been leveraged as a decision support tool to model the series of activities that an organisation performs to deliver a valuable product or service to the market [1], in technical terms, we need to consider the chain that can help extract value from big data in the transport sector. The big data value chain can be used to model the high-level activities that comprise an information system [2]. The chain identifies the following key high-level activities, as shown in the figure below.

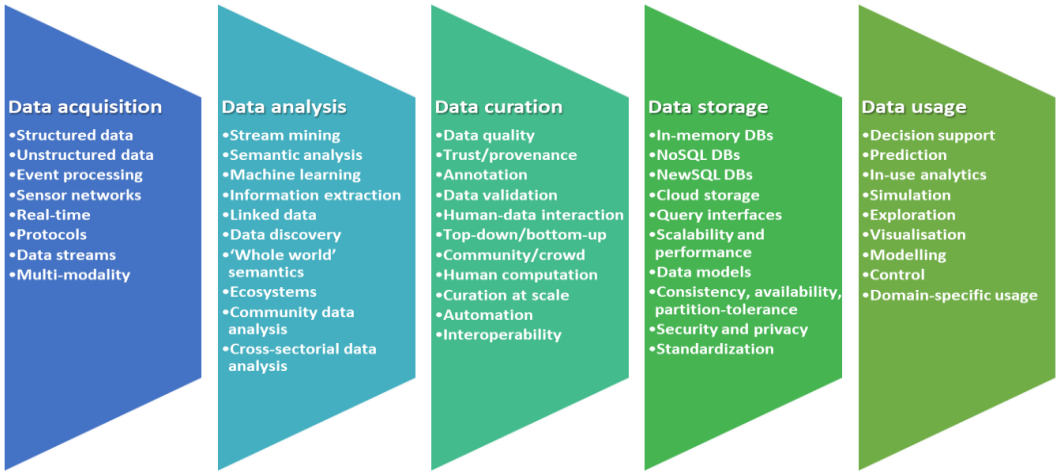


Figure 1 Big data value chain [1]

In LeMO project, we will use **a reference model** described by the above chain to discuss which technical and technological activities related to big data have been used and are appropriate in the transport sector. ***We believe that a more formal guideline can be provided via the well-known reference model on the utilisation of big data.*** According to Curry et al. in [1], each process in this reference model is characterised as follows:

Data Acquisition is the step of gathering, filtering and cleaning data before the data is saved in data warehouses or any other storage solutions on which data analysis can be performed.

Data Analysis is related to making acquired raw data amenable to use for decision-making as well as domain-specific usage.

Data Curation actively manages data over its life cycle to ensure it meets the necessary data quality requirements for its practical usage.

Data Storage is concerned about storing and managing data in a scalable way, satisfying the needs of applications that require access to the data.

Data Usage covers the business goals that need access to data and its analysis, and the tools needed to integrate analysis in business decision-making.

2.2 Socio-economic Impact Analysis (SIA)

Generally, SIA has been used in societal research to analyse environmental impact assessment to identify and assess the potential socio-economic and cultural impacts of a proposed development. If such potential impacts are significant, SIA can assist the developer, and other parties to the process of environmental impact assessment, find ways to diminish, remove or avoid these impacts from happening [3]. To analyse findings and suggestions of LeMO project, we leverage the analysis framework of SIA.

To do that, this section describes the general six steps of SIA as shown in the following Figure 2. In addition, relations of LeMO project with these steps are discussed.

Figure 2 Six steps of SIA [3]

The "Six Steps of SIA" in LeMO project

Scoping: As a preliminary analysis that identified and prioritised SIA considerations and required information, early and adequate scoping narrows the focus of SIA onto issues of potential significance. In this regard, **WP1** of LeMO project belongs to the first step.

Profiling baseline conditions: Focuses on collecting information and context of the proposed development (i.e., on utilisation of big data in LeMO project). The **WPs 1 and 2**, in which we investigated various aspects from the transport sector, are related to this step.

Predicting impacts: Based on the analysis of information from issues scoping and baseline profiling to predict possible socio-economic impacts, it identifies adverse and beneficial issues of a proposed development. In LeMO, WP3 and Tasks 4.1 & 4.2 belong to the stage. In WP3, we conducted a survey and interviews to figure out the issues of utilizing big data in the transport sector, and Task 4.1 focused on barriers and limitations of big data usage. In this task, LeMO will focus on opportunities.

Identifying mitigation: Identifies mitigations including strategies and plans to alleviate, prevent or manage impacts. To do this, LeMO considers the identified barriers and opportunities to provide constructive recommendations and research & policy roadmaps in Tasks 4.3 and 4.4.

Evaluating significance: Involves determining whether a proposed development is likely to cause significant adverse impacts on valued socio-economic components. This step will be carried out in Task 4.5 via a validation workshop in LeMO project. Since a preliminary analysis on the significance of identified barriers and opportunities will be conducted in this deliverable, Task 4.2 also belongs to the step.

Applying mitigation & monitoring: Since application of mitigation measures and efficient monitoring is beyond the objectives and coverage of LeMO project, this step is out of scope. However, LeMO is collaborating with other initiatives and clusters in the EU transport sector as well as relevant projects (e.g., GEKCO), where application of mitigation measures through regulations and policy is being studied. Therefore, efforts will be made to achieve this step through such cooperation.

3 Technological Opportunities

3.1 Big data value chain

The Big Data Value (BDV) reference model (depicted in Figure 3) serves as a reference model to locate Big Data technologies in the overall IT stack [4]. The model is structured in horizontal and vertical concerns.

Horizontal concerns cover specific aspects along the data processing chain. As described in the Big Data Value Chain in Section 2.1, it starts with principles and techniques for data collection and ingestion. Data protection is facilitated by privacy and anonymization mechanisms. This is strongly linked to data management, data processing as well as the area of cybersecurity. Data processing architectures are optimized and scalable for analytics of data. Data analytics provide a better understanding of the data and the meaningfulness of it. On top of it, advanced data visualisation approaches provide enhanced user experience.

Overall it should be noted that the horizontal concerns do not imply a layered architecture. As an example, data visualisation may be applied directly to collected data (the data management aspect) without the need for data processing and analytics [4].

Vertical concerns address cross-cutting issues, which may affect all horizontal concerns. In addition, vertical concerns may also involve non-technical aspects.

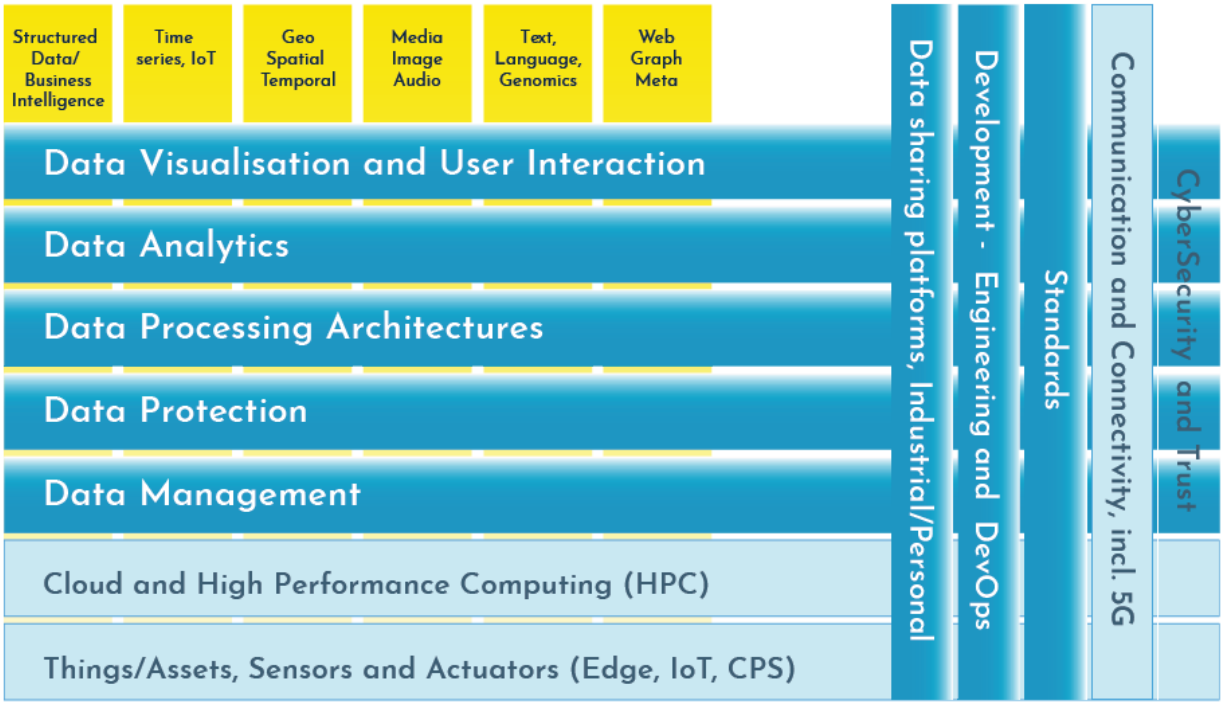


Figure 3 Big Data Reference Model [4]

The following six Big Data types represent the variety of Big Data. They lead to different techniques and mechanisms in the horizontal concerns. (1) Structured data (e.g. Datawarehouse or relational databases); (2) Time series data (e.g. streaming of sensor data); (3) Geospatial data (e.g. GPS data); (4) Media, Image, Video and Audio data; (5) Text data,

including Natural Language Processing data and Genomics representations (e.g. social media data); and (6) Graph data, Network/Web data and Metadata. Standardisations clear the way for fast and easy data integration, exchange and interoperability [4].

The BDV Reference Model is compatible with such standardisations and reference architectures, most notably the emerging ISO JTC1 WG9 Big Data Reference Architecture [5] – now being further developed in ISO JTC1 SC42 Artificial Intelligence [6].

3.2 Big data assessment

The following technological barriers and limitations were mentioned and/or identified in the seven LeMO case studies, in LeMO deliverable D1.1 [7], in NOESIS deliverables 2.1 [8] and 2.2 [2] and the TT pilots [9]–[16]. In this Section, we concentrated on issues, which are (1) caused by or (2) affecting technical relevance. These limitations can be classified into two major groups: These caused by technical restrictions or affecting the technical solution of Big Data in the area of transportation. Furthermore, we provide an evaluation of the severity (1 very low – 5 very high) for every issue listed.

To determine the severity of each limitation, the number of denominations in the different deliverables, the allocation to the corresponding layer of the BDV Reference Model [4] depicted in the below figure, and the six phases of the CRISP-DM are taken into account.

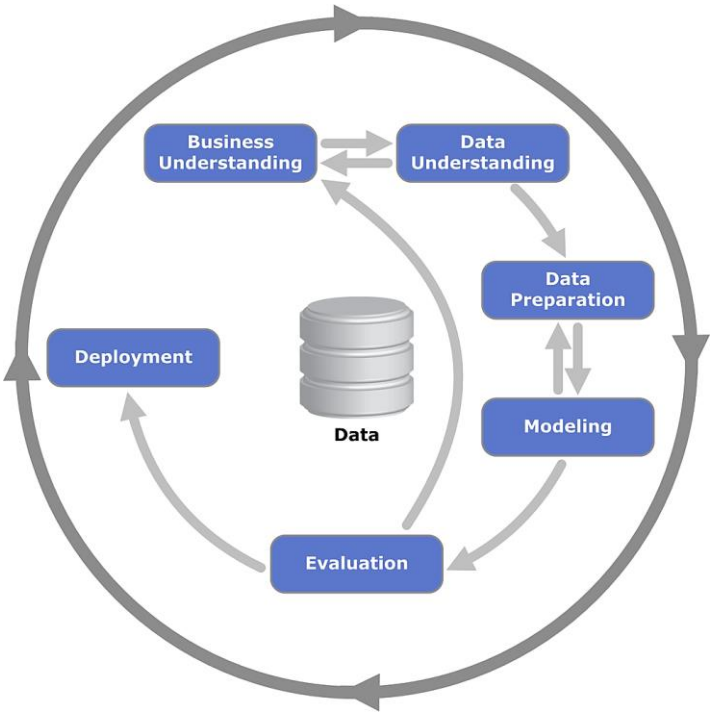


Figure 4 CRISP-DM showing the relationship between the different phases [17]

The CRISP-DM is not a linear process [36], but an iterative process with six phases and numerous bidirectional arrows as shown in Figure 4. The outer circle and the arrows in the diagram

represent the nature of the data mining process. The lessons learned from one phase can trigger another phase, but the process may not end to the stage of Deployment due to the bidirectional dependencies. A detailed description is presented in deliverable D1.3 [18] of the LeMO project.

3.3 Technological opportunities and interventions

3.3.1 Data resource

In deliverable D1.1 [7] we presented four different fields of data which are mainly used in transportation applications: (1) Route-based, (2) vehicle-based, (3) traveller-based, and (4) wide area/external data. All of these fields were used in the analysed LeMO case studies, Transforming Transportation pilots, and NOESIS's use cases. We now focus on the interventions and opportunities for data resources.

Table 1 Interventions and opportunities: Data resource

Code	Description	Source(s)
OPT-TM-DR-1	**Data sources** There are attempts to set up Industrial Data Platforms and Personal Data Platforms (IDPs/PDPs) as marketplaces for providing and sharing data. Furthermore, Research Data Platforms (RDPs) and Urban/City Data Platforms (UDPs) are also available. These platforms facilitate the usage of data but need to be more promoted and supported [4]. Examples are European data portal[3], Mobilitäts Daten Marktplatz[4], or GoData[5]. LeMO CS 2 explored the opportunities and challenges of successfully making Big Data available as open data. And LeMO CS 4 made use of publicly available data to improve their services. *(LIM-TM-DR-1)*	BDVA Strategic Research and Innovation Agenda LeMO CS 2 and 4
OPT-TM-DR-2	**Data quality** There is a general need for the creation of data curation frameworks and workflows. This includes i.e. data curation pipelines, on-line and off-line data filtering techniques, etc [4].	BDVA Strategic Research and Innovation Agenda

[3] https://www.europeandataportal.eu/data/
[4] https://www.mdm-portal.de/about-mdm/technical-details/?lang=en
[5] https://www.govdata.de/

	Furthermore, the data platforms described in OPT-TM-DR-1 would support the provision of quality-assured data. *(LIM-TM-DR-2)*	
OPT-TM-DR-3	**Data exchange** In LeMO case studies 5 and 6 platforms and APIs were used to share transportation relevant information/transport management systems. CS 3 mentioned the advantages of data sharing in cooperative intelligent transport systems. The development of Open-source solutions should be shared and published in the research community and industry. Furthermore, the data platforms described in OPT-TM-DR-1 would support the standardized exchange of data. *(LIM-TM-DR-3)*	LeMO CS 3, 5 and 6 BDVA Strategic Research and Innovation Agenda

3.3.2 Data complexity

In deliverable D1.3 [16], we described the heterogeneity of Big Data technologies in the Hadoop Ecosystem. It follows the classification proposed by Ivanov [18] and made an intuitive representation of a big data platform with the concept of the heterogeneity paradigm. This heterogeneity leads to a complex decision regarding the use of the corresponding technologies and infrastructure, the data collection and processing techniques, and the exploitation of data. But also, some opportunities and interventions to diminish the complexity were identified.

Table 2 Interventions and opportunities: Data complexity

Code	Description	Source(s)
OPT-TM-DC-1	**Choosing the right architecture** Choosing the right architecture is a crucial decision for setting up the fundamentals of a Big Data application. For this reason, extensive benchmarks of systems are carried out and a large number of results are published. There are research approaches like Aloja[6] and DataBench[7] to support decision-makers by extracting information on benchmark runs on different hardware and software configurations. This information supports Big Data	Aloja [21], DataBench [22] LeMO CS 1

[6] http://aloja.bsc.es/
[7] https://www.databench.eu/

	practitioners to choose the corresponding architecture and technologies. LeMO CS 1 presented an already configured Big Data platform as a (cloud) service for customers. This can be customized for the individual needs of the customer. *(LIM-TM-DC-1)*	
OPT-TM-DC-2	**Data collection and processing** The active development of data collection and processing solutions. In LeMO CS 4 the data comes in their provider's data centre through an API network and then it goes to a number of different outlets. The data is normalised into internal proprietary formats using Google buffers. In LeMO CS 1 a cloud platform for data collection, Stream Processing, Machine Learning framework, and Data Acquisition was presented. These solutions should be shared along with research communities and industry. BDVA states that to achieve the agility demanded by real-time business and next-generation applications, a new set of interconnected data management capabilities is required [4]. There are also emerging standardization activities to support the implementation in industry [5]. *(LIM-TM-DC-2 and LIM-TM-DR-3)*	LeMO CS 1, 4 BDVA Strategic Research and Innovation Agenda, ISO/IEC TR 20547-2:2018
OPT-TM-DC-3	**Exploiting data** Understanding and analysing the great variety of data is crucial and at the same time one of the biggest challenges of Big Data applications. Therefore, there are several areas where research is still needed: • Semantic and knowledge-based analysis • Content validation • Analytics frameworks and processing • Advanced business analytics and intelligence • Predictive and prescriptive analytics • High Performance Data Analytics (HPDA) • Data analytics and Artificial Intelligence This can lead to an improved usage of data and the information which needs to be extracted [4].	BDVA Strategic Research and Innovation Agenda LeMO CS 1, 3, 4, 5, 6 and 7

	Providers of LeMO case studies 1, 3, 4, 5, 6, and 7 developed their own analytics and exploitation strategies to make use of the data available to them. It could give a chance to share their know-how and strategies for the same purpose in the transport sector *(LIM-TM-DR-2 and LIM-TM-DC-3)*	

3.3.3 Limitations of the used infrastructures and systems

Also, some **limitations of the used systems** were mentioned with respect to the challenges of using big data technologies and the huge amount of data. This resulted in limitations of the existing infrastructure as well as the limitation of the existing traffic management systems to handle the great amount of data collected by various sources. Some LeMO case studies already outlined solutions to deal with these problems:

Table 3 Interventions and opportunities: Limitations of the used infrastructures and systems

Code	Description	Source(s)
OPT-TM-LI-1	**Limitations of the existing infrastructure** In LeMO CS 1 a cloud infrastructure was presented, which is customizable to the needs of the customer. Cloud infrastructures provide a great variety of services that can be used like infrastructure, platform, or software as a service. *(LIM-TM-LI-1 and LIM-TM-DC-1)*	LeMO CS 1
OPT-TM-LS-1	**Limitation of existing traffic management systems** LeMO CS 5 presented a solution to overcome the restrictions of the existing traffic management system. A new system had to be constructed from scratch to provide all necessary functionalities and technologies. *(LIM-TM-LI-2)*	LeMO CS 5

3.3.4 Affecting technical solutions

Also mentioned in the referenced sources are limitations which are (1) caused by or (2) affect technical relevance. These limitations can be classified into two major groups: Those caused by technical restrictions or those affecting the technical solution of big data in the area of transportation. Interventions or opportunities affecting these are the following:

Table 4 Interventions and opportunities: Affecting technical solutions

Code	Description	Source(s)

OPT-TM-AT-1	**Lack of skilled workers** The company introduced in LeMO CS 6 developed its application by hiring only Freelancers. In addition, there are European initiatives to fill the gap of skilled workers like EDSA[8], EDISON[9], and more and more data science student programs [4]. Furthermore, there are Big Data solutions that are provided as services. A lot of the complexity is reduced by the service provider so that the customer can focus on their business case (e.g. LeMO CS 1). *(LIM-TM-AT-1)*	LeMO CS 1 and 6 BDVA Strategic Research and Innovation Agenda
OPT-TM-AT-2	**High investment costs** In addition, most of the technology used in the Big Data ecosystem is open source. Thus, the investment costs result from setting up and running the hardware as well as the development and management of the applications. LeMO CS presented a cloud infrastructure that can be used by their customers. This diminishes the initial costs of the infrastructure. *(LIM-TM-AT-2)*	LeMO CS 1
OPT-TM-AT-3	**Security, Privacy and Trust** In LeMO CS 4 the blockchain technology was used to increase transparency and trust for importers and exporters by replacing the need for extra time and personnel to record transactions within the supply chain and benefiting shippers to minimize information disruption, reduce paperwork needed, and maintain access to original information without worrying about tampering during and after shipment. There are also various Horizon 2020 projects as PAPAYA [10] addressing the tension between data analytics of sensitive data and security, privacy and trust concerns. Also, the results of more general projects of privacy and security in cloud applications as PRISMACLOUD[11] and PaaSword[12] can be considered.	LeMO CS 4 PAPAYA [23], PRISMA CLOUD [24], PaaSword [25]
OPT-TM-AT-4	**Heterogeneity of individuals**	LeMO CS 6,

[8] http://edsa-project.eu/
[9] https://cordis.europa.eu/project/rcn/198292/factsheet/en
[10] https://www.papaya-project.eu
[11] https://prismacloud.eu/
[12] https://paasword.io/

	LeMO CS 6 presented the application of big data technology and analysis methods to provide a multimodal public transport information and route planning service to passengers. This highlights the opportunities of Big Data applications as they can provide the individuals the possibility to choose between different alternatives. Technology can even support the heterogeneity of people by making use of (machine learning algorithms in) recommender systems that take the preferences of each customer into account. [20] *(LIM-TM-AT-4)*	Recommender Systems [20]
OPT-TM-AT-5	**Multiplicity of applications** LeMO CS 6 presented a solution on how to combine the data of different companies and data providers to create a single point of information on a specific topic. One strength of Big Data applications is to merge different kinds of data sources and to extract specific information from it. *(LIM-TM-AT-6)*	LeMO CS 6
OPT-TM-AT-6	**Lack of business models** There is a large number of different business models to make use of the opportunities that the use of Big Data technology provides. In LeMO CS 4 advanced predictive analytics technologies were used to build a real-time logistics tool to increase the efficiency of delivery vehicles. Furthermore, there are European research and innovation projects, like DataBench[13], which try to close the gap between the performance of business KPIs and the used Big Data infrastructure and technology. In general, there are three keyways to generate value for companies along the Big Data value chain, regardless of sector or domain: • Optimizing and improving the core business; • selling data services; • and, perhaps most importantly, creating entirely new business models and business development. Identifying these business models will be an important challenge [4]. *(LIM-TM-AT-7)*	LeMO CS 4, DataBench [22] BDVA Strategic Research and Innovation Agenda

[13] https://www.databench.eu/

4 Policy and Legal Opportunities

4.1 Policy issues

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

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

On such basis, the table below lists three core opportunities related to public and private policies based on the preliminary results included in D1.2. As for the public policies, additional opportunities can be found under Section 4.2 related to the legal issues. The barriers are discussed in D4.1.

Table 5 Interventions and opportunities: Public and private policies

Code	Description	Source(s)
OPT-POL-PU-1	There is a general tendency towards data openness and data sharing in public policies across the EU, including in the transport sector.	New (inspired by LeMO D1.2 and D4.1)
OPT-POL-PU-2	The current policy framework has helped improve the protection of individuals (e.g. data subjects, consumers, etc.). This provides fertile ground to have individuals adopt new technologies.	New
OPT-POL-PU-3	The current (recently updated) policy framework has contributed to the sharing of data between companies, and between companies and governments.	New

OPT-POL-PU-4	The current policy framework is characterized by a drive towards harmonization of various legislative frameworks across the EU Member States. This has created a better level playing field for companies wishing to deploy cross-border data-driven applications and technologies. *(LIM-POL-PU-1)*	New
OPT-POL-PU-5	Recent changes to the applicable policy framework have eliminated certain geographical barriers (e.g. free flow of non-personal data) and thus provide fertile ground for cross-border data-driven applications and technologies.	New
OPT-POL-PU-6	The adoption of soft law in relation to data, such as in particular in the context of data sharing, has positively triggered a change in mentality among private actors on the market. *(LIM-POL-PR-2)*	New

4.2 Legal issues

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

In the below diagram, we provide an overview of the core 13 legal issues identified in the context of Deliverable D2.2:

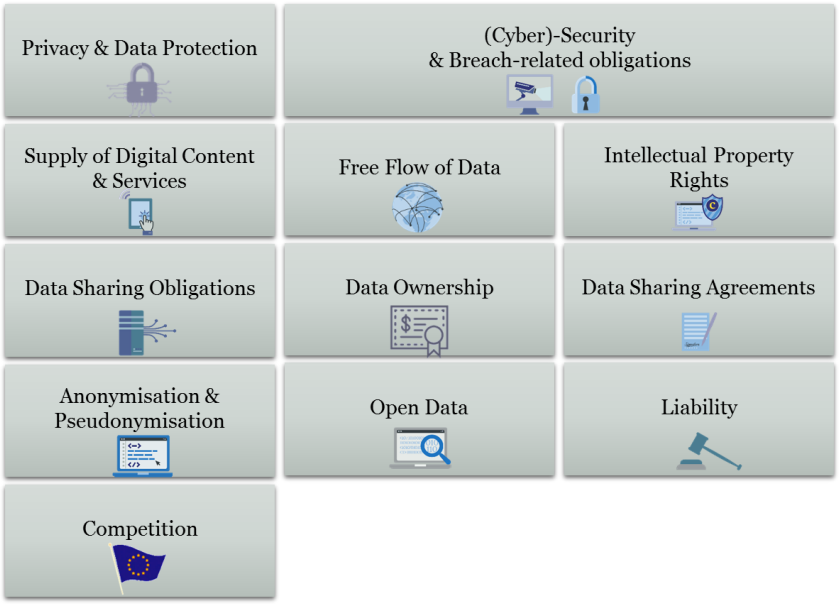


Figure 5 Overview of 13 legal issues identified in Deliverable D2.2

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

4.2.1 Privacy and data protection

Certain principles and requirements related to privacy and data protection can be difficult to fit with some of the main characteristics of big data analytics. In this respect, Deliverable D2.2 demonstrated that finding a balance between the various interests at stake is of paramount importance. In light hereof, it is essential to keep in mind that the right to the protection of personal data is not an absolute right but must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality.

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

Table 6 Interventions and opportunities: Privacy and data protection

Code	Description	Source(s)
OPT-LEG-DP-1	Having well-managed, up-to-date and relevant data may help improve data quality. This is also an opportunity to comply with the GDPR which requires implementing measures to disregard the elements of a database that would be inaccurate. Ultimately this is an opportunity to improve the data management and contribute to a better analytics outcome.	LeMO D2.2
OPT-LEG-DP-2	The development of coordinated and EU-wide guidance and templates, taking into account complex data processing activities, at EU and national level will likely increase legal certainty for those involved in the data value chain, and ultimately benefit data subjects. *(LIM-LEG-DP-1; LIM-LEG-DP-5; LIM-LEG-DP-6; LIM-LEG-AP-7; LIM-LEG-FF-2)*	LeMO D2.2
OPT-LEG-DP-3	The legal requirements, and in particular the requirement to maintain records of personal data processing activities and the requirement to perform Data Protection Impact Assessments and Legitimate Interests Assessments, provide the opportunity to carefully examine the privacy implications of big data analytics and to identify the best measures to be implemented for compliance with the GDPR, such as minimisation of data collection, transparency towards data subjects, providing control mechanisms to data subjects (e.g.	LeMO D2.2

	opt-out), etc. *(LIM-LEG-DP-7; LIM-LEG-DP-10; LIM-LEG-DP-12)*	
OPT-LEG-DP-4	Big data allows building and developing new consent models and provides more and novel automation, both in the collection and withdrawal of consent. *(LIM-LEG-DP-12)*	LeMO D2.2
OPT-LEG-DP-5	The requirements of "data protection by design" and "data protection by default" will require changes within organisations in order to adopt new approaches in the development of processes, services and products. These new obligations can be an opportunity for stakeholders in the data value chain to improve their offering by integrating or further developing privacy-enhancing technologies and solutions, and ultimately to comply with many other requirements of the GDPR. *(LIM-LEG-DP-3; LIM-LEG-DP-7; LIM-LEG-DP-10)*	LeMO D2.2
OPT-LEG-DP-6	Although considering the rights of data subjects and anticipating their concrete application can be difficult in a technology-rich environment, new technologies can also be seen as an opportunity to allow individuals to exercise their rights as innovative and responsible engineering can facilitate, among others, the exercise of individuals' rights of access, objection, withdrawal of consent, rectification, as well as data portability. *(LIM-LEG-DP-10; LIM-LEG-DP-12)*	LeMO D2.2
OPT-LEG-DP-7	The new general accountability principle laid down under the GDPR has provided the opportunity to organisations to demonstrate their compliance with the many data protection obligations included in the GDPR and has thus increased organisations' responsibility towards end-users in relation to privacy and data protection.	LeMO CS 6
OPT-LEG-DP-8	The novel privacy and personal data protection requirements in the EU have increased harmonisation across the EU, thus improving legal certainty. This has notably improved the situation of cross-border projects, products and services (including the development of new data-driven technologies).	New
OPT-LEG-DP-9	Actors of the data value chain, including authorities, standardisation bodies, service providers, vendors and industry players can develop together standards, certification mechanisms, seals, marks and codes of conduct, which can	New (inspired by LeMO D2.2)

	improve accountability and compliance with the GDPR. *(LIM-LEG-DP-1)*	
OPT-LEG-DP-10	Organisations' improved compliance with the GDPR gives the opportunity to increase end-users' trust in their products and services, as well as to give them a competitive advantage over less-compliant and/or less privacy-preserving organisations, products or services.	New

4.2.2 (Cyber-)Security and breach-related obligations

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

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

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

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

Table 7 Interventions and opportunities: (Cyber-)Security and breach-related obligations

Code	Description	Source(s)

OPT-LEG-CBO-1	Big data analytics can contribute to the detection of security issues, and thus help comply with certain core obligations included in various legislations such as the GDPR or the NIS Directive.	LeMO D2.2
OPT-LEG-CBO-2	The cyber-security requirements can ultimately help creating a cyber resilient culture across organisations, notably giving the opportunity to train and certify staff.	LeMO D2.2
OPT-LEG-CBO-3	Actors of the data value chain, including authorities, standardisation bodies, service providers, vendors and industry players can develop together standards, certification mechanisms, seals, marks and codes of conduct, which can improve compliance with various legislations such as the GDPR or the NIS Directive. *(LIM-LEG-CBO-4; LIM-LEG-CBO-5; LIM-LEG-CBO-9)*	LeMO D2.2
OPT-LEG-CBO-4	The heightened security requirements permit to enhance supply chain trust and resilience by engaging third party suppliers and customers in cybersecurity processes and business continuity measures.	LeMO D2.2
OPT-LEG-CBO-5	The new (or improved) incident notification requirements, including those under the GDPR and the NIS Directive, provide the opportunity to actors of the data value chain (including in the transport sector) to better tackle security incidents, in a quicker, more efficient and concerted manner.	New

4.2.3 Anonymisation and pseudonymisation

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

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

Table 8 Interventions and opportunities: Anonymisation and pseudonymisation

Code	Description	Source(s)
OPT-LEG-AP-1	Irreversibly anonymised personal data may be processed without the need to comply with data protection laws.	LeMO D2.2 LeMO CS 2 and 3
OPT-LEG-AP-2	The application of anonymisation techniques may engender an exemption from the notification obligations related to personal data breach. *(LIM-LEG-CBO-9)*	LeMO D2.2
OPT-LEG-AP-3	Anonymisation techniques may serve as a means to comply with data protection law, and specifically with the following obligations: (i) data protection by design and by default; (ii) security of processing; (iii) purpose limitation; and (iv) storage limitation. *(LIM-LEG-DP-7; LIM-LEG-DP-8; LIM-LEG-DP-10; LIM-LEG-CBO-4)*	LeMO D2.2 NOESIS D4.1 LeMO CS 6
OPT-LEG-AP-4	Anonymisation techniques may prove to be apt instruments to protect non-personal information in a technical manner. If successful, this may encourage stakeholders involved in the big data value cycle to engage in data sharing. *(LIM-POL-PU-2)*	LeMO D2.2

4.2.4 Supply of digital content and services

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

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

Table 9 Interventions and opportunities: Supply of digital content and services

Code	Description	Source(s)
OPT-LEG-DIG-1	The new legal framework on the supply of digital content and services introduces a high level of protection for	LeMO D2.2

	consumers paying for a service but also for those providing (personal) data in exchange for such service. The provision of digital content and services whereby the consumer provides his or her personal data requires the trader to take into account the underlying economic reality of transactions using personal data. This ultimately protects individuals with regard to the processing of their personal data.	

4.2.5 Free flow of data

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

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

Table 10 Interventions and opportunities: Free flow of data

Code	Description	Source(s)
OPT-LEG-FF-1	The Free Flow Regulation eliminates certain geographical barriers to EU data flows and thus improves the situation for cross-border data-driven applications and technologies.	NOESIS D4.1
OPT-LEG-FF-2	Following the introduction of the Free Flow Regulation, stakeholders expect significant cost reductions for cloud storage and processing, necessary for big data analytics services in the transport sector.	LeMO D2.2
OPT-LEG-FF-3	The elimination of data localisation requirements will lead to increased competition. This in turn will create more innovation, which will positively impact big data analytics in the transport sector.	LeMO D2.2

OPT-LEG-FF-4	Following the introduction of the Free Flow Regulation, it will be easier for SMEs and start-ups to access new markets when data localisation requirements are eliminated.	LeMO D2.2

4.2.6 Intellectual property rights

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

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

Table 11 Interventions and opportunities: Intellectual property rights

Code	Description	Source(s)
OPT-LEG-IP-1	If stakeholders in the big data analytics lifecycle are able to rely on intellectual property rights to protect their investment (be it monetary or intellectual) in (parts of) the datasets, they may be more eager to engage in data sharing in a big data analytics context, including in the transport sector. *(LIM-POL-PU-2)*	LeMO D2.2

4.2.7 Open data

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

information sharing has not been made mandatory for public undertakings (yet), the new regime constitutes a significant development for the transport sector, where services are often provided by public undertakings.

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

Table 12 Interventions and opportunities: Open data

Code	Description	Source(s)
OPT-LEG-OD-1	Open data can help eliminate barriers to market entry for start-ups and SMEs, by giving them access to resources which they would not be able to access otherwise.	LeMO D2.2 LeMO CS 4 and 7
OPT-LEG-OD-2	Open data may lead to an increase in competition and opens up potential for private companies to innovate, which creates value for the economy.	LeMO D2.2 LeMO CS 4 and 5
OPT-LEG-OD-3	Open data can create public-private sector collaborations and synergies, which in turn create economic value and are another ground for innovation.	LeMO D2.2

4.2.8 Data sharing obligations

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

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

Table 13 Interventions and opportunities: Data sharing obligations

Code	Description	Source(s)
OPT-LEG-DS-1	Data sharing obligations imposed through legislation may offer opportunities for increased competition and innovation by opening up data to private actors which would otherwise not have access to such data.	LeMO D2.2

OPT-LEG-DS-2	Typically, when data must be shared with other private actors, we see that some kind of remuneration may be demanded, allowing the businesses involved to recover the related costs.	LeMO D2.2
OPT-LEG-DS-3	The rise of technical specifications and standardisation requirements, which are often necessary to ensure interoperability in the context of Intelligent Transport Systems, could also lead to an increased adoption of data sharing requirements.	New
OPT-LEG-DS-4	Public tenders could be an alternative means of imposing data sharing obligations. *(LIM-POL-PR-2)*	LeMO D2.2

4.2.9 Data ownership

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

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

Given the nature of the issue, no opportunity in relation to data ownership has been identified.

4.2.10 Data sharing agreements

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

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

Table 14 Interventions and opportunities: Data sharing agreements

Code	Description	Source(s)
OPT-LEG-DSA-1	At the time of writing, data sharing agreements provide the only solution to govern access to and/or exchange of data between the numerous stakeholders active in the big data value cycle, including in the transport sector.	LeMO D2.2
OPT-LEG-DSA-2	The contractual freedom left to the parties when drafting data sharing and related agreements allows them to shape their arrangements as they see fit on the basis of the many possible varying circumstances.	New

4.2.11 Liability

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

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

Table 15 Interventions and opportunities: Liability

Code	Description	Source(s)
OPT-LEG-LI-1	The ongoing work of the EU institutions regarding extra-contractual and statutory liability specifically, and the envisaged non-regulatory and regulatory interventions, may be beneficial for the uptake of big data in the transport sector and improve legal certainty in relation thereto. *(LIM-LEG-LI-2)*	LeMO D2.2

4.2.12 Competition

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

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

Table 16 Interventions and opportunities: Competition

Code	Description	Source(s)
OPT-LEG-COM-1	The opportunity to acquire data-rich companies should not cause competition issues under merger control regimes. What matters is the kind of data you are acquiring, how unique it is, whether it can be easily replicated and whether you can shut out rivals. *(LIM-LEG-COM-3)*	LeMO D2.2
OPT-LEG-COM-2	The transport sector has always collected and analysed large quantities of data, such as data from timetables, traffic news and air schedules. Big data allows this to be used to create more efficient and smarter transport systems for people and freight and increases the scope to monetise and sell data for new and innovative services.	LeMO D2.2
OPT-LEG-COM-3	New app developer opportunities such as short journey planning for multiple modes of transport in major cities by collating open data feeds, real-time traffic information based on crowd-sourced data from smartphones and vehicle GPS data, real-time public transport journey planning by combining public transport data with information crowd-sourced from users through smartphones, suggested driving routes based on traffic information crowd-sourced from users through smartphones.	LeMO D2.2

5 Ethical and Social Opportunities

This section introduces interventions to diminish the barriers and limitations, which we identified in D4.1 for ethical and social aspects. In addition to the interventions, opportunities related to these aspects are discussed in this section.

5.1 Trust

As revealed in CS2 of LeMO project, low quality of data is caused by duplication or diverse errors of data which increase with intervention or participation of many actors in distributed environments. And it could harm data trust. The quality might be lower when using Big Data without appropriate data standards. Therefore, standardisation has been done by authorised organisations and EU in transport fields to reduce impacts of heterogeneous data formats and to encourage to utilise uniform standards. In this regard, one of the best ways to keep data trust is cooperation between various actors such as governmental organisations and relevant companies, as mentioned in many cases such as CS2, CS3, CS5 and CS6 of LeMO project. Meanwhile, related companies gain the right to access other useful data following the standards, and its utilisation makes the big transport data more valuable in terms of trust aspect. These activities generate an ecosystem in which the virtuous circle keeps data trust. And such an ecosystem would be the best or the most achievable practice in order to keep data trust in the transport sector. Other important trust issue is providing service to users, as considered in all three projects. The improvement of service quality increases customer satisfaction. The quality of service is closely related to reliability, comfort, accessibility, convenience, connectivity, openness, quality control & monitoring and waiting/service time. Such elements would be achieved by using Big Data analytics, as shown in CS6 (i.e., optimised transport & improved customer service) of LeMO project and Pilot 6 and 7 (i.e., integrated urban mobility and dynamic supply chain networks) of TT project. For example, customer satisfaction obtained from various data (e.g., social media data) can be a key aspect to be monitored in order to understand the trust level of services. Moreover, in the aforementioned ecosystem, relatively small companies could find an opportunity to increase the trust level of their service via participating in standards utilisation.

Table 17 Interventions and opportunities: Trust

Code	Description	Source(s)
OPT-ES-TR-1	Validation via data aggregation and cross-analysis of various data could improve the low quality of data *(LIM-ES-TR-1 and 2)*.	LeMO CS 2, 3, 5 and 6 NOESIS and TT (Many cases)
OPT-ES-TR-2	In order to improve reliability, comfort, accessibility, convenience, connectivity, openness, quality control &	LeMO CS 2, 3, 5, 6 and 7

	monitoring and waiting/service time, big data analytics will be used in transport sector.	
OPT-ES-TR-3	Organise an ecosystem in which various actors cooperate with each other to use big transport data and encouraging to join in the ecosystem that uses official standards. It will help to alleviate gaps about trust of services between big and small companies.	LeMO CS 6 and 7
OPT-ES-TR-4	Higher volume in big data leads to other chances for trustworthiness, such as reputation.	LeMO D2.3

5.2 Surveillance

Since the surveillance using CCTV has privacy issues, one of the safest ways is to avoid its installation in places where any personal information could be included. For example, in CS2, relevant authorities check whether a place and direction could include any personal properties or not. However, it is not possible in most cases, or it becomes useless in terms of surveillance objectives. Alternatively, strictly regulated access permissions according to levels of surveillance data, in order to reduce misuses, as in CS3, can be implemented. Unfortunately, it still carries risks regarding usage of Big Data such as identifying a person via analysing aggregated data. It could be more severe in rural areas, where the number of public transport users is relatively low. Therefore, there is a need to introduce a threshold system, where data for cities below a certain activity level threshold is not made public, as suggested in CS6.

Table 18 Interventions and opportunities: Surveillance

Code	Description	Source(s)
OPT-ES-SUR-1	There would be no direct solution for barriers and limitations in terms of the surveillance issue since this issue always relates to privacy issue. Instead, avoiding installation in places, where any personal information could be included, and strictly regulating access permissions according to levels of surveillance data to reduce misuses could alleviate serious concerns of the surveillance issue *(LIM-ES-SUR-1)*.	LeMO CS 2, 3 and 6 NOESIS (4 use cases)
OPT-ES-SUR-2	A threshold system where data for cities below a certain activity level threshold is not made public could be adopted to reduce the threats about identifying a specific person via big data analysis *(LIM-ES-SUR-2)*.	LeMO CS 3, 6 NOESIS
OPT-ES-SUR-3	Improve transport safety via surveillance when needed by selecting essential places and assuring closed utilisation.	LeMO CS 3

OPT-ES-SUR-4	As the degree of sophistication increases, the centralised management of data, such as traffic data, needs to be enabled, and combined with optimisation of massive data usage. If this is successful, a high degree of precious data can be gathered to support decision-making. This way, big data can facilitate smart transport. In other words, it can encourage reliability; availability; maintainability; safety; but also, efficiency.	LeMO D2.3

5.3 Privacy

Relevant actors who collect photos and videos from CCTV on roads and publish the data to create added value could face specific privacy-related issues. Since these data could have personal information, this kind of data should be carefully checked whether the data includes personal properties and information or not, as CS2 and CS3. In addition, as shown in CS5, data such as AIS-data are classified as personal data. The data which are only available for traffic management purposes can be possessed by participants anyway. However, they are not available as open data. It is related to assigning access permission aforementioned in trust aspect. Contrary agreements could be made between relevant stakeholders and government about the situations when the government should use this type of data. In this regard, there is a trade-off between protecting privacy and extending usage of big transport data. The relation could be somewhat controlled by introducing data sensitivity for privacy issue, as proposed in CS6 of LeMO project.

Table 19 Interventions and opportunities: Privacy

Code	Description	Source(s)
OPT-ES-PRI-1	There are technical solutions to reduce the privacy problem due to data including personal information. For example, classifying personal level of data, protecting data by encryption and adopting data sensitivity for trade-off between privacy and utilising data *(LIM-ES-PRI-1)*.	LeMO CS 2, 3 and 6 NOESIS
OPT-ES-PRI-2	In the transport sector, consideration for not only personal data but also data from vehicles is essential. To ensure the privacy of individual vehicle data, allowing authorised actors to access and use the data might be one of the effective solutions *(LIM-ES-PRI-2)*.	LeMO CS 3 and 5 NOESIS
OPT-ES-PRI-3	In certain situations, individuals may be willing to forsake part of their privacy in return for the benefits that big data applications offer.	LeMO D2.3

OPT-ES-PRI-4	If organisations involved in big data analytics are able to address the users' privacy needs, this could open the door to more people engaging in big data.	LeMO D2.3

5.4 Free will

Consideration of free will is necessary, when data is related to people or their mobility. In transport sector, low quality of traffic information could lead to threats of free will and loss of various resources. For example, an improper decision about traffic restrictions based on analysis of traffic data will waste time resources and return with large amounts of claims. It decreases users trust for the service. This situation expected by CS providers is one of the factors hindering the adoption of advanced technologies of traffic management in real industry. However, free will might not be violated as final decision is up to people. Because suggested choices do not force people and just help them to avoid searching and analysing huge data. Moreover, free will is a bit neglected for specific purposes like in CS5, if the objectives are not related to people. It enables relevant stakeholders to save resources and be more punctual.

Table 20 Interventions and opportunities: Free will

Code	Description	Source(s)
OPT-ES-FW-1	Free will could be threatened by the low quality of transportation information. Adopting learning data model to avoid wrong decisions in future might be useful to deal with the issue *(LIM-ES-FW-1)*.	LeMO CS 5
OPT-ES-FW-2	Though high technology suggesting options could seem like they limit free will, big data analytics could support more freedom by providing explanations of the choices, such as recommending travel paths with explanations.	LeMO CS 2 and 5 D2.3
OPT-ES-FW-3	Social and political phenomena can be extended into new domains by big data, as it achieves greater accuracy in pinpointing individual behaviour and can be used as a supporter for free will.	LeMO D2.3
OPT-ES-FW-4	Increasing accessibility and personalisation for people may provide proper choices through a better understanding and serving people's needs.	LeMO D2.3

Data ownership: Open data seems to be free from data ownership by agreements between stakeholders or interventions of governments, as shown in CS2 and CS5. According to the

investigation in CS2 of LeMO, trying to extend range of open data in transport sector by states in EU shows an optimistic view. However, if it is for commercial reasons, data ownership should be considered to avoid any confusion of owning original data, as shown in CS2, CS3 and CS7. Moreover, it is also related to user right when data include any personal information. In this case, based on agreements of users, the personal user data should only be operated within the necessary processes as shown in CS7.

Table 21 Interventions and opportunities: Data ownership

Code	Description	Source(s)
OPT-ES-DO-1	To reduce the conflicts about data ownership between commercial activities, make clarification of ownership via adopting advanced techniques such as blockchain *(LIM-ES-DO-1)*.	LeMO CS 2 and 7 NOESIS (11 use cases)
OPT-ES-DO-2	To avoid threats for user rights of personal data ownership, it should be strictly regulated to ask users' agreement and to use the data within the necessary processes, since it is also related to the privacy issue *(LIM-ES-DO-2)*.	LeMO CS 3 and 5 NOESIS
OPT-ES-DO-3	Given the nature of the issue, no opportunity in relation to personal data ownership has been identified. We believe however that the opportunities related to control, discussed above, may (at least partly) address the data subjects' concerns with respect to the "ownership" of their personal data.	LeMO D2.3

5.5 Social discrimination

Although there is side-effect such as losing workplaces by introducing more efficient strategy based on Big Data analytics as CS1, most cases had no social discrimination issues that may occur while using Big Data in transport sector.

Table 22 Interventions and opportunities: Social discrimination

Code	Description	Source(s)
OPT-ES-SD-1	To reduce discrimination levels for offering and utilising information between users and companies, personalised services including explanations could be provided. It should encourage the use of public data that can diminish relevant issues *(LIM-ES-SD-1)*.	LeMO CS 1

OPT-ES-SD-2	Big data analytics can be a tool to make existing discriminatory decisions visible.	LeMO D2.3
OPT-ES-SD-3	Big data analytics can be used to tailor customised services that meet the needs of certain social groups in order to improve their inclusion in society.	LeMO D2.3

5.6 Environment and others

As discussed in CS3, CS5 and CS6, transport sector must consider environmental issues such as emission of gases, particles and noises from roads and traffics. The importance of efficient traffic management increases with growing urbanisation. In many cases of LeMO, NOESIS and TT projects, utilisation of Big Data has positive influence on environment, such as reduction of pollution emissions, animal run-over, noise and vibration and decarbonisation. In addition, data services seem not to have a close effect on environment as shown in CS2 and CS4. However, there might be rebound effects of an increased use of Big Data in transport sector. According to report D2.4 of LeMO, it leads to increased volumes of ICT-infrastructure in which more smart devices (e.g. IoT devices), sensors, data centres and higher capacity of networks are needed to acquire, store and process data. For example, in one of LeMO case studies, route planning applications will have to tackle the challenge of creating bottlenecks by guiding all commuters to the same optimal path. If this is not addressed then it may lead to a rebound effect, contributing to further congestion instead of addressing it. As shown in CS6, such pitfalls can be avoided through real-time tracking of traffic flows and by creating limits that prevent the same path from being suggested to commuters after a threshold is reached. Like, other cases, in which there might not be negative effects, environmental issues should be considered as much as possible. For example, as debated in CS7 of LeMO, mobility on demand services should consider environmental aspects and do not replace existing public transport offers that are already more environmentally friendly.

Table 23 Interventions and opportunities: Environment

Code	Description	Source(s)
OPT-ES-ENV-1	Reduction of pollution emissions, animal run-over, noise and vibration and decarbonisation could be achieved by optimisation of traffic management and suggestion of appropriate decisions via big data analytics *(LIM-ES-ENV-1)*.	LeMO CS 5 and 6 NOESIS and TT (Many cases)
OPT-ES-ENV-2	Consideration of any potential environmental issues and harmonisation with existing public transport offers that are already more environmentally friendly can alleviate rebound effects due to increase of ICT infrastructure.	LeMO D2.4

Table 24 Interventions and opportunities: Others (i.e., Transparency, consent and control)

Code	Description	Source(s)
OPT-ES-OT-1	**Transparency:** Providing transparent information to individuals whose personal data is involved in big data analytics may increase trust in the processing activities and the technology used. When people feel they can trust technology, they tend to be more willing to engage in it *(LIM-ES-OT-1)*.	LeMO D2.3
OPT-ES-OT-2	**Consent:** Both industry and government should take up the responsibility to eliminate the misconceptions that exist regarding personal data protection. Data subjects should be educated, notably through transparent notices from industry, about the grounds for processing and the possible impacts on privacy *(LIM-ES-OT-2)*.	LeMO D2.3
OPT-ES-OT-3	**Control:** Giving control to data subjects should not necessarily stifle the use of big data. Instead, a bigger involvement of data subjects may lead to improved analytics given that the data subjects can correct mistakes and detect unfair decisions *(LIM-ES-OT-3)*.	LeMO D2.3

6 Economic and Political Opportunities

In this section, we will discuss the opportunities and interventions that can be undertaken by different actors to support the big-data transport business model in being successful.

6.1 Focus area 1: Value creation from Big Data

There are many opportunities afforded by big data and the data economy. Most of these have been discussed in D2.1 in Section 3.1 and 3.2 that describe the potential benefits to the transport system and transport actors to act as data suppliers. These are summarized in Table 25.

Table 25 Opportunities from big data in the transport sector

Direct improvements to the transport system	Benefits of transport sector actor as data supplier
1. Shared situational awareness **2. Improving transport network capacity** • Long term planning of the transport network • Realtime traffic management • Mitigating risks that degrade network capacity **3. Improvement of transport services** • Marketing activities in the transport sector • Vehicle routing • Service provision and schedules **4. Shift towards sustainable mobility and transport**	**1. Data for profit** **2. Exchange data for "other" data useful in their business model** **3. Support complementary services providers with data** **4. Support research in the research industry to potentially benefit from them** **5. Promote public/club goods – i.e. altruism or corporate social responsibility programs.**

Note that often in order to promote public or club goods this may sometimes be imposed by public authorities or other "club" members or managers. Providing data allows other members to benefit, though it may not be directly beneficial for the business model of data supplier. Non-compliance may result in penalties or expulsion from the club.

The case studies also identified benefits of big data that validated the general list above. The opportunities and benefits identified in D3.2 are presented below.

Table 26 Opportunities and benefits afforded by big data identified in the case studies

Case study	Opportunities and benefits
CS 1: "Railway transport"	Improved predictive maintenance leads to different types of direct and indirect benefits for the rail company and train service provider. • More punctual and reliable train services.

	• Cost decrease of train services for passengers and logistic operations. • Reduction of greenhouse emission in the transport sector. • Higher competitiveness of rail to other modes like air travel. • European-wide cooperation of involved rail operators and other stakeholders.
CS 2: "Open data and the transport sector"	Open data policies and supporting technologies and services: • Help in creating and extending new data services, products and markets. For example, to support other transport planners to improve their awareness and consider the road conditions and public transport service availability. • Authorities are actively collaborating with other organisations with experience and resources to handle such data technologies. Thus, pushing forward data technologies and practices. • Improved insight in the transport network for performance and safety.
CS 3: "Real-time traffic management"	Improve real-time traffic management • Increasing real-time information availability can support seamless travel. • Big data used to recreate high quality reproduction of current traffic and traveller conditions, but also optimising the future traffic situation. Reaching precision in models much higher than before. • Better-organized traffic systems mean a better flow of vehicles on the road, and it means no idling cars, buses, and trucks in traffic jams. This may translate to shorter operation time, optimized utilization of fuel, and lower pollution.
CS 4: "Logistics and consumer preferences"	Use of sensitive data from vehicles in route optimization • Big data predictive analytics solutions (such as the LiveRoad Geospatial analytics platform) applied to road and transport to forecast risks and delays. • Cooperation with various public stakeholders can support traffic safety aims. • Successful collaboration between private and public sector increases future collaboration necessary for smart city initiatives.
CS 5: "Smart inland shipping"	Real-time data on the status of bridges and berths available in ports • Information that helps skippers and road users better plan their journey and arrival, with less hassle and annoyance and without unnecessary fuel consumption and exhaust emissions. • Helps route optimisation by delivering real-time information of bridges. This limits the time road vehicles are waiting for bridges. Data on the

	availability of berth places limits the skipper search time for a place to moor and saves fuel. • Public-private sector projects lay an inter-organisational collaboration platform. The participants have gained support for the project within their organisations and have built trust and procedures for further cooperation. Together with ambition to further develop the platform, this is fertile soil for more cooperation. With the foundation being there, the entry costs for new participants are low. Experience, know-how and best practices are available for other governmental authorities to profit from. • Open data for anyone, enabling public and private parties to use the data for innovations.
CS 6: "Optimised transport & improved customer service"	Smart mobility and routing solutions for public transport. • Integrated data set across different operators that is also open for integration of other data sources. • Begins transformation of a very conventional industry to work together to offer various improved services and value to customer.
CS 7: "Big data and intelligent transport systems"	On-demand mobility services that provide and rely on big data • Connect ex-urban, rural or not well-connected areas with existing public transport system; • Decrease barriers to public transport access via offering first and last mile mobility services; • Simplify payment procedures by offering integrated ticketing systems for different MOD service providers; • Improving urban living conditions, like pollution, noise, space, by reducing number of privately-owned vehicles and the urban traffic; • Reducing greenhouse emissions by implementing more efficient and environmentally friendly mobility offers like electrically operated car sharing fleets or extended public transport; • Improving first and last mile goods and packages delivery in urban areas with intelligent and centralized MOD services, like using cargo-bikes, centralized distribution-hubs, etc.

The opportunities for value creation in transport are summarized in Table 28**.**

Table 27 Opportunities for value creation with big data in transport

Code	Description	Source(s)
OPT-EP-VC-1	**Shared situational awareness.** Increasing shared awareness of the state of transport activity (and other information) between **n**etwork managers, transport operators (managers	LeMO CS 1, 2, 3, 5 and 7

	to drivers), customers, public authorities. The ability of Big Data Analytic (BDA) to collect various data streams and to create meaningful information out of it is the basis of this particular opportunity.	
OPT-EP-VC-2	**Long term planning of the transport network.** BDA supports transport network/ infrastructure and services planning in the long term. Conventional models can be enhanced, and new BDA techniques can be developed to create more data-driven approaches.	LeMO CS 3 and 4
OPT-EP-VC-3	**Real-time traffic management**. High velocity data collection, transmission and analysis can support real-time control of each transport actor, especially in network management. Note that these refer also to access to nodes, thus it may support warehouse planning, port capacity, etc.	LeMO CS 3, 5 and 7
OPT-EP-VC-4	**Mitigating risks that degrade network capacity.** Some of the key risks that BDA may help to detect well in advance and therefore help to prevent via maintenance, are in track, vehicle failure or adverse weather conditions.	LeMO CS 1 and 4
OPT-EP-VC-5	**Marketing activities in the transport sector.** Collection of data about potential customers support service providers in understanding the customers better and therefore providing more targeted services. This is a key step in any service or infrastructure plan.	LeMO CS 2, 3 and 7
OPT-EP-VC-6	**Vehicle routing.** Vehicle use can be optimized by using a variety of data, especially that which may indicate disruptions or lower speed in the area (e.g. weather/road conditions and congestion).	LeMO CS 3, 4, 5 and 6
OPT-EP-VC-7	**Service provision and schedules.** Service planning (similar to network infrastructure planning) can be done in a more precise way that conventionally. It can also be performed closer to real-time to react to the changing market.	LeMO CS 2, 4, 6 and 7
OPT-EP-VC-8	**Reduce cost of transport services.** Several actions, such as improved maintenance, optimized routing, and better network management, can reduce the costs of providing transport services. This can lead to better value for customers, in the short and long term. Important in this respect is a proper cost benefit assessment calculation	New LeMO CS 1, 3 and 4

	especially considering the cost of investment and operating of big data technology.	
OPT-EP-VC-9	**Shift towards sustainable mobility and transport.** Optimized services can reduce externalities (e.g. pollution, greenhouse gas emissions, noise) and also support public transport services and sustainable services (using electric vehicles) that are overall better for the environment.	LeMO CS 1, 3, 5, 6 and 7
OPT-EP-VC-10	**Data for profit.** Data that is produced while carrying out the activities may be sold to other companies/sectors for additional income. This drives innovation, specifically in data collection. E.g. is the collection of data that improves weather prediction and **nowcasting.**	LeMO CS 4
OPT-EP-VC-11	**Exchange data for "other" data useful in their business model.** Network managers especially will profit from sharing data about their network performance in order to get data from their transport network users. This can be used to create better prediction models of transport demand.	LeMO CS 1
OPT-EP-VC-12	**Support complementary services providers with data.** Data can be used to improve the transport service ecosystem of other non-competitors. This is part of the drive for airports to share open data, to empower other service providers to step in and indirectly make the airport better (i.e. by providing last mile services.)	CS 2, 5 and 7
OPT-EP-VC-13	**Support research in the research industry to potentially benefit from them.** Researchers often need industry data to create better methods. Sharing data can be a means to create synergies between the transport and research industry. Other benefits may also come from the collaboration that benefit the transport industry in their adoption of BDA.	LeMO CS 1, 2, 3 and 4
OPT-EP-VC-15	**Promote public goods – i.e. altruism or corporate social responsibility programs.** Open data can support other companies without transaction cost to companies that will provide good services to the public.	LeMO CS 2

6.2 Focus area 2: Data technology and systems

Data technology – referring to the technical systems and infrastructure needed for access, storage, transmission and analytics - is still in development. Technology develops together with research and use in businesses. Research needs on data technology from the technological perspective is discussed in detail in Section 3.

Businesses seeking to develop and integrate big data technology in their operations will also need to carry out experiments. A suitable form of experimentation is the living lab. This deals with several different issues that have been highlighted in D4.1.

- **EP-IR-1**: Bias from data sample and algorithms.
- **EP-IR-2**: Bad quality of data.
- **EP-IU-1**: Technology is still maturing.

A living lab is composed of two key aspects: (1) a test environment embedded in a real-world operation system, and (2) support for cyclical development [28]. In the transport sector, this will require the support and different degrees of participation from different stakeholders. Private actors and users, i.e. those part of the business ecosystem, must be willing to implement the experiments involving the new technology and absorb the risks that are inherent in the prototypes. Knowledge institutes, such as research institutes or universities also play a role in research and development. This may also be carried out in-house if the company is sufficiently large and has sufficient competence to also "improve" the technology. Finally, public actors may support in terms of funding, temporary permissions, and as clients to transport services.

Table 28 Interventions and opportunities: Data technology and systems

Code	Description	Source(s)
OPT-EP-DTS-1	**Living labs on the development and use of data technology.** Technology develops together with research and use in businesses. A living lab is composed of two key aspects: (1) a test environment embedded in a real-world operation system, and (2) support for cyclical development. (LIM-EP-IR-1; LIM-EP-IR-2; LIM-EP-IU-1)	New

6.3 Focus area 3: Data availability

Data availability is influenced by two general issues. (1) Poor data quality and (2) Restricted data access. Both can be dealt with quite similarly at the firm level. The research and policy interventions have two aims:

- Remove, replace and augment data sources
- Circumvent restrictions to data access

6.3.1 Remove, replace and augment data sources

Data management processes are needed to identify problematic data sources and the best course of action to deal with it. Part of the business process is to decide if the "bad" data source implies that the company will need to pause or stop the operations depending on the data. This could be the case for highly sensitive operations or technical machinery, e.g. in aircrafts or other transport infrastructure.

Firms can consider fixing the data sources. For instance, this might involve fixing or replacing the data collection systems, such as ITS systems, vehicle sensors, and traffic cameras. Older systems could be replaced with newer ones that produce better data quality, more suitable for BDA, i.e. 5Vs. If sources are external to the company, contractual best practices could be considered to ensure that data quality is enforced as part of the contractual best practices (see Section 4.2.10 Data sharing agreements). Further research might be necessary to understand how best to monitor data quality.

Firms might also consider augmenting the data sources, i.e. modify how data analytics is performed. Multiple data sources could be used, which increases the robustness of the analysis and reduces reliance on the lower quality data source. Augmented ML might also be researched to supplement machine learning with human input.

6.3.2 Circumvent restrictions to data access

There are several reasons why data is not or should not be "accessible" to a firm. Technological solutions could be applied, for instance to comply with privacy and legal restrictions (see D2.2 on anonymization/pseudonymisation or remote working, i.e. access to data only through restricted servers). The industry might also consider implementing data standards that support data sharing with complex permissions and licensing systems.

There are several initiatives currently underway. As with many standards and frameworks critical mass is necessary to legitimize the practice in the industry. The International Data Spaces Association has developed a reference architecture model and certification for software that "allows for modelling, configuring, monitoring and enforcement of the rules and polices specified in data contracts[14]". With the right technology, various business models can be developed for different data sharing needs and specifications.

Table 29 Interventions and opportunities: Data availability

Code	Description	Source(s)
OPT-EP-DA-1	**Improving data sources**. For instance, this might involve fixing or replacing the data collection systems, such as ITS	LeMO D4.1

[14] International Data Spaces Association, JOINTLY PAVING THE WAY FOR A DATA DRIVEN DIGITISATION OF EUROPEAN INDUSTRY, https://www.internationaldataspaces.org/wp-content/uploads/2019/10/IDSA-strategic-paper-for-EU.pdf

	systems, vehicle sensors, and traffic cameras. Older systems could be replaced with newer ones that produce better data quality, more suitable for BDA, i.e. 5Vs. If sources are external to the company, contractual best practices could be considered to ensure that data quality is enforced as part of the contractual best practices. Further research might be necessary to understand how best to monitor data quality. (LIM-EP-IR-1; LIM-EP-IR-2)	LeMO CS 4 LeMO D2.2
OPT-EP-DA-2	**Augmenting the data sources.** i.e. modify how data analytics is performed. Multiple data sources could be used, which increases the robustness of the analysis and reduces reliance on the lower quality data source. Augmented ML might also be researched to supplement machine learning with human input. (LIM-EP-IR-1; LIM-EP-IR-2; LIM-EP-RD-1; LIM-EP-RD-2; LIM-EP-RD-3)	New
OPT-EP-DA-3	**Adopting standards and best practices for data sharing with the right access controls**. Encouraging the industry to adopt standards happens gradually if the standards have concrete value. Research to develop the right kind of standards still needed. (LIM-EP-RD-1; LIM-EP-RD-2; LIM-EP-OC-3; LIM-EP-OC-5)	LeMO D4.1

6.4 Focus area 4: Organizational challenges

Transport companies that are integrating big data analytics into their processes need to deal with the current lack of expertise and competence within their own companies and perhaps even in the sector. There are several risks involved with the poor management of data and the analytics, as described in previous deliverables, thus requiring good data scientists and workers. Furthermore, in some cases, the change brought about by integrating big data in the company, especially to make decisions at the strategic level, is the need for change management. Several options are discussed below.

6.4.1 Enhance the labour market

A company has multiple ways to increase human resource capacity if it currently lacks the competence. Gig economy platforms such as Fiverr or WeWorkRemotely[15] might help to gain short term experts from the non-local labour market. However, teleworking options have their disadvantages and risks and should be implemented cautiously. Labour rates might be slightly

[15] WeWorkRemotely, https://weworkremotely.com/

higher than an employee, quality assurance procedures may not be currently suited to remote workers and freelancers, and continuity and loyalty for long term projects. While technological means have been developed to enable remote working on the same physical systems (i.e. data storage servers, analytics, etc.), data scientists need domain knowledge and close cooperation with operations and the business functions that are being transformed. In many cases, especially in the exploration of potential, regular communication and interaction within the team is necessary. Hence, good teleconferencing technology and other co-creation platforms would also be a good research opportunity [29].

Immigration policy can also be simplified and modified to attract foreign talent (or migrated citizens) from outside the EU. Estonia has promoted the E-Resident program aimed at enabling foreign talent to start companies serving the EU market[16]. The UK has a special visa "Tech Nation Visa Application - Tier 1 (Exceptional Talent) Visa for Digital Technology" where those who are "exceptionally talented or promising individual in the digital technology sector" can apply[17].

Developing *home-grown* talent is important for the long-term viability. An understanding of the key skills needed in the future should guide the training programs designed now. The SFIA Foundation has developed a framework, the Skills Framework for the Information Age, for the types of skills that are needed to support the digital age "the globally accepted common language for the skills and competencies related to information and communication technologies, digital transformation and software engineering." [18]. Some of the relevant key themes are cyber security, digital transformation, big data and informatics. Universities and the industry/government could work together to develop integrated skills focused programs. Further, the funding for research projects that further extends the current data science would help to attract top educators/researchers to sustain skills development.

6.4.2 Support change management

Support for change management may in many ways already be attainable. Funds should be channelled to support the transport industry to know where and how to get the help they need. This might apply to both those using BDA and those profiting from BDA-driven transport services. Public authorities and third parties could support change management via three approaches:

1. Guidelines
2. Tools and methods, e.g. software or ROI calculation sheets, evaluation frameworks
3. Financial incentives – ("Business expansion: Where a company extends its business activities (for example, insourcing, development of new products or services, addressing new markets or client groups, including internationalisation)"[19].

[16] E-Resident program, https://e-resident.gov.ee/

[17] TECH NATION VISA, https://tech-nation-visa.smapply.io/prog/tech_nation_visa_application/

[18] SFIA FOUNDATION, https://www.sfia-online.org/en/framework/sfia-7/about-sfia-7

[19] Public policy and support for restructuring in SMEs, https://www.eurofound.europa.eu/observatories/emcc/comparative-information/public-policy-and-support-for-restructuring-in-smes

Each may require research for a more specific analysis of requirements and best practices.

6.4.3 Encourage start-ups and spin-offs

Organizations that find it costly and difficult to reinvent their organization may consider instead supporting spin offs or start-up like ventures. The investment to begin the organization might be better served than retraining older workers. If successful, the organization could be reintegrated into the company or allowed to operate independently. Public funding to support start-ups could also be made available from public funds, which support entrepreneurship and social causes. The spinoffs could be oriented to supplementary business models and new ways to use or collect data, such as led to the creation of loki (See Case study 7 in [1]).

Table 30 Interventions and opportunities: Organizational challenges

Code	Description	Source(s)
OPT-EP-OC-1	**Support the use of remote working and short-term employment.** Gig economy platforms can help to fill the gap in personnel in the very short term, but new methods of securely, effectively and creatively working together must be developed and researched. (LIM-EP-OC-4)	New
OPT-EP-OC-2	**Immigration policy to be simplified and modified to attract foreign talent (or migrated citizens) from outside the EU.** Different visas with different incentives could be proposed to ensure that those with specific skillsets or have started businesses needed to boost the sector and investment in the EU. (LIM-EP-OC-4)	New
OPT-EP-OC-3	**Develop training and education programs ready for the future skills demand.** Use global skills framework like SFIA to guide training programs. Industries and universities should work together to develop relevant industry driven training programs. Research funding made available to attract researchers to academia and to develop data science research clusters. (LIM-EP-OC-2; LIM-EP-OC-4)	LeMO D2.2
OPT-EP-OC-4	**Support change management.** Organizations need help to coordinate the organizational structure and processes to effectively and efficiently make use of BDA driven insights. Public authorities and third parties could support change management via three approaches: Guidelines, Tools and methods, and financial incentives. (LIM-EP-OC-3; LIM-EP-OC-5)	New

OPT-EP-OC-5	**Encourage start-ups and spin-offs.** Organizations that find it costly and difficult to reinvent their organization may consider instead supporting spin offs or start-up like ventures. If successful, the organization could be reintegrated into the company or allowed to operate independently. Funding to support start-ups could also be made available from public funds, which support entrepreneurship and social causes. (LIM-EP-OC-3; LIM-EP-IU-3)	New

6.5 Focus area 5: Business risks and costs

6.5.1 Circumventing high investment costs for infrastructure

While the full costs of investment for big data infrastructure is high, there exists already services that offer Infrastructure-as-a-Service models for storing and even processing big data with the flexibility of different technological and scale options [30]. This makes it easy to experiment, explore and start with big data technology. ROI calculation still remains an important way of assessing make-or-buy decisions, especially if costs and benefits are well-known (See also Section 3 Technological opportunities under “Limitations of the used infrastructures and systems”).

6.5.2 Understanding costs and risk of data

Guidelines and assessment frameworks could also be created to support the identification of risks in the holding of data. Storage of data is a cost, not just on the operations of storage servers, but also in the terms of risks that so-called dark data brings, which are more of a business and legal risk. Any company collecting data or experimenting with IoT should be made aware of this and know how best to handle situation.

6.5.3 Preparing for cybersecurity risks

Actors dealing with data need to have structures in place to deal cybersecurity attacks or malpractice. The level of investment is usually commensurate with the severity of the threat (i.e. the impacts.) However, this requires a threats assessment and perhaps standardized audits conducted by third party certification companies. The NIST Cybersecurity Framework might be appropriated for European companies. Relevant questions regarding the awareness that a transport operator using BDA and holding of big data constitutes a level of risk that needs to be assessed.

Table 31 Interventions and opportunities: Business risk and cost

Code	Description	Source(s)
OPT-EP-BRC-1	**Use Infrastructure-as-a-Service models to avoid ownership costs of big data infrastructure.** While the full costs of investment for big data infrastructure is high, there exist already services that offer Infrastructure-as-a-Service models for storing and even processing big data with the flexibility of different technological and scale options. This makes it easy to experiment, explore and start with big data technology. ROI calculation still remains an important way of assessing make-or-buy decisions, especially if costs and benefits are well-known. Cloud storage and cloud computing are key service trends, which have proven to have a good business model. (LIM-EP-IU-1; LIM-EP-IU-2)	LeMO D4.1, D2.1
OPT-EP-BRC-2	**Guidelines and assessment frameworks could also be created to support the identification of risks in the holding of data.** Storage of data is a cost, not just on the operations of storage servers, but also in terms of risks that so-called dark data brings, which are more of a business and legal risk. Any company collecting data or experimenting with IoT should be made aware of this and know how best to handle such situations. (LIM-EP-OC-1; LIM-EP-OC-5)	New
OPT-EP-BRC-3	**Preparing for cybersecurity risks.** Actors dealing with data need to have structures in place to deal with cybersecurity attacks or malpractice. The level of investment is usually commensurate with the severity of the threat (i.e. the impacts.) However, this requires a threats assessment and perhaps standardized audits conducted by third party certification companies. The NIST Cybersecurity Framework might be appropriated for European companies. Relevant questions regarding the awareness that a transport operator using BDA and holding of big data constitutes a level of risk that needs to be assessed. (addressed in Section 4.2.2.) (LIM-EP-OC-1)	LeMO D2.2

6.6 Focus area 6: Sustainability in business models

BDA applications in the transport sector may have external social and environmental impacts. Sustainability should be firmly integrated into the business models and the institutional context.

6.6.1 Responsible data analytics

Ethical guidelines and data science practices should be researched, developed and adopted throughout the industry. Two different complementary practices could be taken. First, to support the integrating of ethics in data science itself, as the MIT example provides[20]. Research into the topic of machine ethics is important as it is perceived as the necessary step for a broader acceptance and use of the "machines" in transport operations[21].)

Second, actors should increase the transparency of what they are developing, using the outcomes of their actions (to the extent that they know). Transparency can be supported using *watchdogs* that can support the monitoring, assessment and troubleshooting. This is one of the strategies currently employed for social media network giant, Facebook[22,23]. While such practices could easily amount to simply whitewashing practices, independent third-party watchdogs with transparent methods are also being used to ensure sustainability practices, such as Carbon Dioxide reporting in transport[24].

6.6.2 Assessment of social impact

New technology, especially as the outcome of the ever-expansive digital revolution, often requires rethinking of societal values and new norms in practice. While ethics and norms develop and react to disruptions often naturally, continuous dialogue is needed to ensure that policy directives as well as products and services currently being developed are flexible to anticipate potential issues. These new norms should be made clear, where known. These new norms should also influence government assessment procedures. For instance, the way checks and balances work in the government and in civil society might need to be updated in light of the capability of BDA systems to make quick assessments and highlight issues in the transport system. Recently, there have also been discussions on the appropriate method to assess public project funding, conventionally done using the Cost-Benefit Assessment [31]. BDA might enable new methods to carry out the assessments. However, it might also require other values to be included, just as climate change impact is also regularly included in assessments. The shifts must be at the paradigm level that updates the conventional criteria for distribution of economic welfare to something else.

[20] https://shass.mit.edu/news/news-2019-ethical-ai-design-interview-abby-everett-jaques

[21] https://www.nature.com/news/machine-ethics-the-robot-s-dilemma-1.17881

[22] https://www.cjr.org/the_media_today/facebook-fact-checking.php

[23] http://ec.europa.eu/information_society/newsroom/image/document/2019-5/facebook_baseline_report_on_implementation_of_the_code_of_practice_on_disinformation_CF161D11-9A54-3E27-65D58168CAC40050_56991.pdf

[24] https://www.smartfreightcentre.org/en/how-to-implement-items/what-is-glec-framework/58/

6.6.3 Maintaining integrity of public institutions

Public authorities have a special role in society to maintain a fair system and to ensure the implementation of systems for the public welfare. Hence, the government as it implements the policies of elected representatives must remain responsible and in control of the means used [32]. In the era of BDA, where processes might not be sufficiently transparent or clear to non-data science experts, additional checks must be carried out to also ensure the results are acceptable. Governance structures might need to be modified to institute additional checks. Governments might increase reliance on different types of external experts, from the commercial or university sector. These must also be monitored to ensure that the provided services are in line with the government values. Frameworks to develop policies could be used to ensure integrity of new policies and governmental actions[25].

Table 32 Interventions and opportunities: Sustainability in Business Models

Code	Description	Source(s)
OPT-EP-SBM-1	**Develop and integrate ethics and virtue in analytical systems and machines.** Research into methods to achieve this should be carried out and encouraged in the transport industry. (LIM-EP-IR-1, LIM-EP-PR-1)	LeMO D2.3
OPT-EP-SBM-2	**Rely on external watchdogs to increase transparency.** Actors should increase the transparency of what they are developing, using the outcomes of their actions (to the extent that they know). Transparency can be supported using watchdogs that can assist in the monitoring, assessment and troubleshooting of the company. (LIM-EP-IR-1, LIM-EP-PR-1)	New
OPT-EP-SBM-3	**Develop new ethical assessment frameworks to supplement existing economic-based assessments. I**t might also require other values to be included, just as climate change impact is also regularly included in assessments. The shifts must be at the paradigm level that updates the conventional criteria for distribution of economic welfare to something else. This is required at the policy level to ensure that the new policy relying on BDA has a thorough understanding of negative (and positive) impacts and how these should be balanced at the societal level. (LIM-EP-IR-1, LIM-EP-PR-1)	LeMO D2.3

[25] https://www.bigpolicycanvas.eu/

7 Socio-Economic Impact Analysis

7.1 Analysis purpose and methods

In this section, a Socio-economic Impact Analysis (SIA) is conducted to deeply understand the findings from D4.1 [19] and this deliverable. To avoid confusion in our analysis, a definition of the term "impact" in our SIA is required. Impacts are potential changes caused – directly or indirectly, in whole or in part, for better or for worse – by development activities [33]. Based on this definition of 'Mackenzie Valley Environmental Impact Review Board', we define the term "impact" in our SIA as follows: ***Impacts mean the intended or unintended consequences, that could result from different factors such as barriers/limitations on one hand and opportunities/interventions on the other***.

In view of this, we perform an analysis to gain insights on how different aspects (technological, policy & legal, ethical & social; and economic & political) may be impacted through the combined influence of barriers/limitations on one hand and opportunities/interventions on the other and what this could mean for the use of big data in transport sector, as considered and investigated during the LeMO project. Table 33 lists some of the questions considered during this analysis and describes the methods used to answer these questions. With awareness of their limitations in this particular analysis, Natural Language Processing (NLP) techniques have also been used and their outcomes have been corroborated through comparative readings of the tables and discussions. *Note that we will not discuss results obtained by applying NLP in order to avoid limiting potential and possible interpretations of the results. Instead, we display all the results in the following sections and Appendix B of this deliverable to share it with others and describe how the results have been derived to provide the justification of the used methodology.*

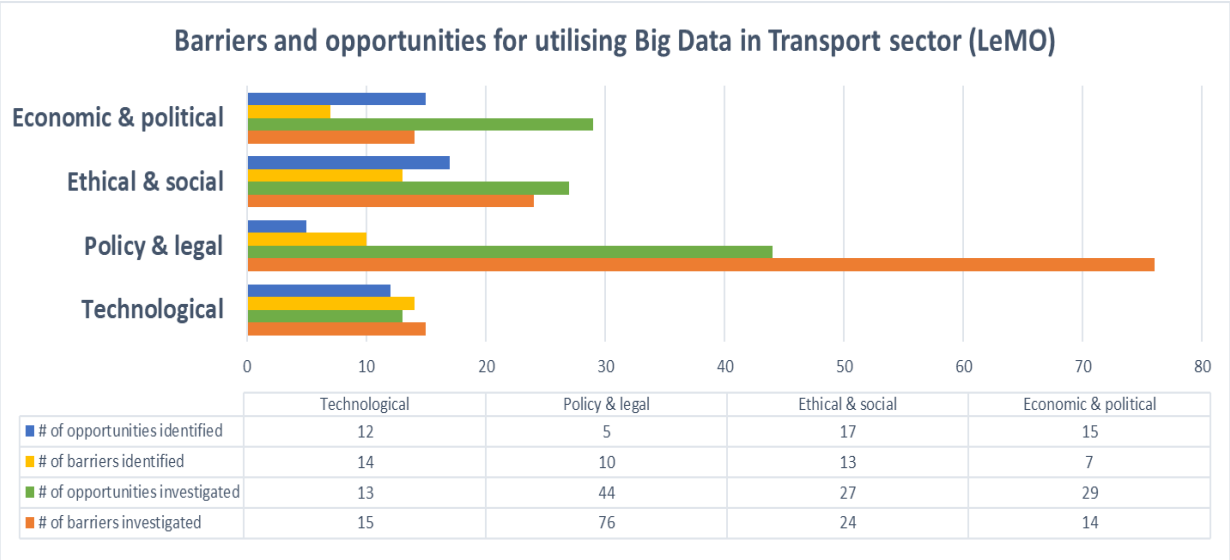
Table 33 Questions and methods for Socio-economic Impact Analysis

Question	Method for answering the query
1. If the number of barriers/limitations listed for each aspect is bigger than the number of opportunities/interventions listed for that aspect or vice versa?	Answered through a comparative reading of the tables in D4.1 [19] and this deliverable.
2. In which aspects are there more barriers/limitations and in which are there less?	Answered through a comparative reading of tables in D4.1.
3.In which aspects are there more opportunities/interventions and in which are there less?	Answered through a comparative reading of tables in D4.2.

4. Do the Barriers/limitations and Opportunities/interventions interact with each other across different aspects?	Answered through a combination of comparative reading of tables as well as NLP techniques.
5. How strong is this interaction?	Answered through a comparative reading of tables, NLP techniques as well as discussions among partners/stakeholders.
6. Whether this interaction reinforces or cancels the effect of the barriers/limitations and opportunities/interventions?	Answered through a comparative reading of tables, NLP techniques as well as discussions among partners/stakeholders.
7. How do we compute severities of barriers/limitations and intensities of opportunities/interventions?	Answered through NLP techniques

7.2 Analysis based on comparing item numbers

In this section, we compare item numbers for each aspect to answer questions 1, 2 and 3 in the Table 33. The following chart summarises the item numbers of each aspect in terms of barriers and opportunities. The items are categorised into investigated and identified groups. Investigation means that we have induced the items by literature review, and identification stands for founding the items in LeMO's case studies. NOESIS's use cases and TT's pilot cases fall among the investigated items.



	Technological	Policy & legal	Ethical & social	Economic & political
# of opportunities identified	12	5	17	15
# of barriers identified	14	10	13	7
# of opportunities investigated	13	44	27	29
# of barriers investigated	15	76	24	14

Figure 6 Item numbers of barriers and opportunities

As visible from the Figure 6, compared to other aspects, the economic & political aspect has more items of opportunities/interventions (O/I) than that of barriers/limitations (B/L). This is

because the use of big data in the transport field provides a lot of occasions for creating value, leading to higher number of O/I. It also results from the fact that many researchers and practitioners trying to apply big data in the transport sector. For the ethical and social aspect, the O/I are higher than the B/L, since practitioners are expecting that the use of big data in the transport sector would lead to more personalized and efficient services thus reducing social discrimination and addressing environmental concerns.

Conversely, for the policy & legal aspect, the B/L are much higher than O/I. This is because the regulatory framework for the use of big data is very complex and ambiguous. Also, it does not offer many opportunities and may need strong interventions to foster the use of big data in the transport sector.

Lastly, the number of B/L is only slightly higher than O/I for the technological aspect. This is because though there may be several technical complications, there is also a range of steps and measures available to overcome these limitations.

Moreover, we can also compare the numbers of B/L with that of O/I for each issue or category, as shown in the Figure 7.

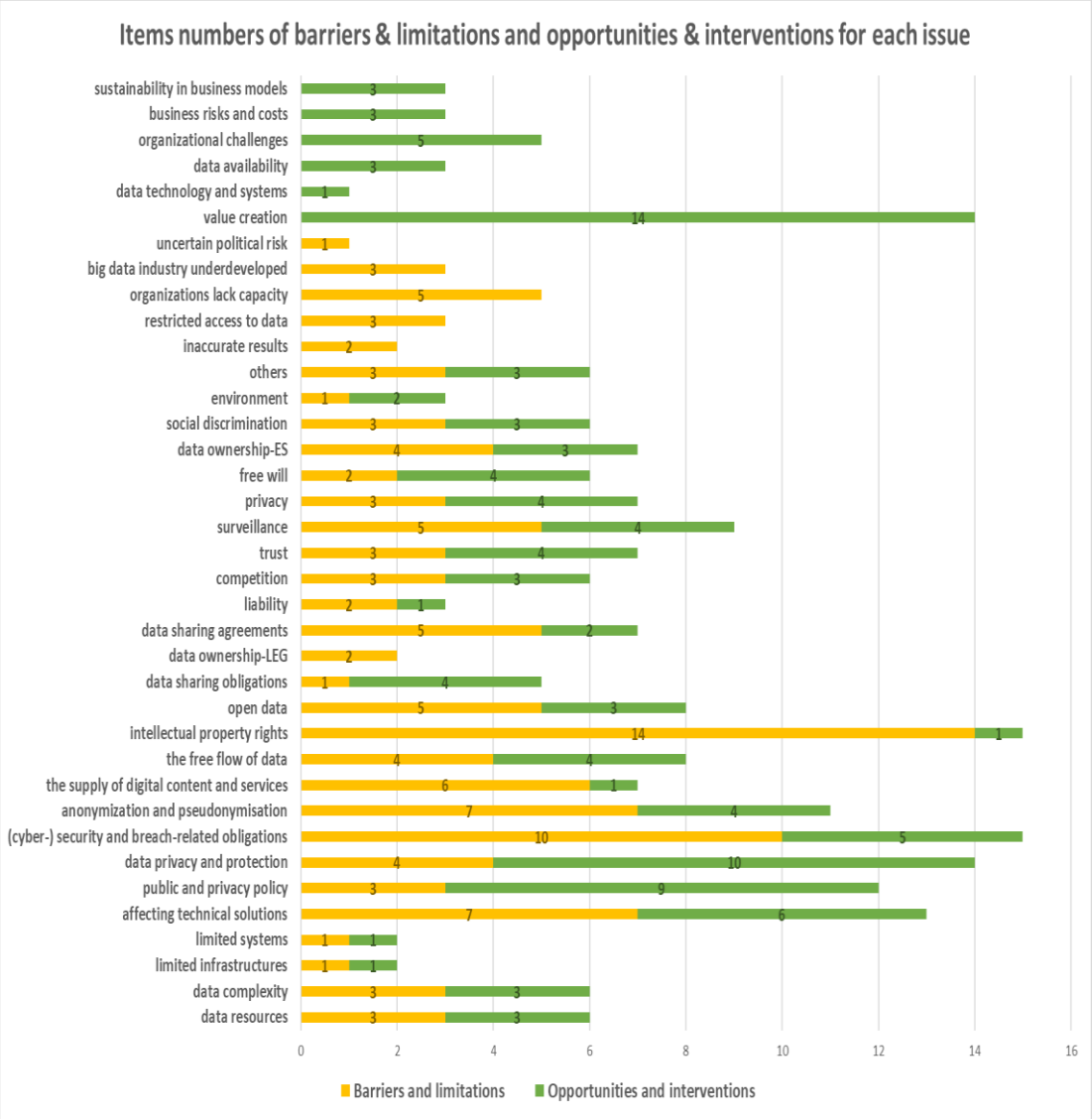


Figure 7 Comparison of item numbers for each issue

7.3 Severity and intensity analysis by LNP and visualisation techniques

7.3.1 Dataset for SIA in LeMO

In this section, we describe the dataset used for our SIA in LeMO project. We have generated the dataset by collecting descriptions of items [i.e., barriers/limitations (B/L) and opportunities/interventions (O/I)] in tables of D4.1 and this deliverable. Each B/L and O/I was represented by a specific code, (for example, OPT-EP-SBM-#) for the ease of statistical analysis, and the item means each component for the code (i.e., OPT-EP-SBM-1 or 2). Category indicates sub-groups of issues. The following Table 34 summarises statistical information of the dataset.

Table 34 Statistical information of the dataset for SIA

Property	Value (explanation)
Total number of words in dataset	10,626
Total number of items (i.e., B/L and O/I)	243 (B/L: 129, O/I: 114)
Average number of words in the description of each item	43.728 (minimum: 5, maximum: 128)
Number of aspects considered	4 aspects
Average number of items for each aspect	30.375 (minimum: 14, maximum: 76)
Average number of issues or categories containing different items (i.e., B/L and O/L) for each aspect	7.875 (minimum: 5, maximum: 14)
Average number of items for each issue	3.857 (minimum: 1, maximum: 14)

The following diagram shows relations between items and aspects for B/L and O/I. Circles made by dots represent B/L and circles made with lines represent O/I. Circles made with both lines and dots represent items that were mentioned as both B/L and O/I.

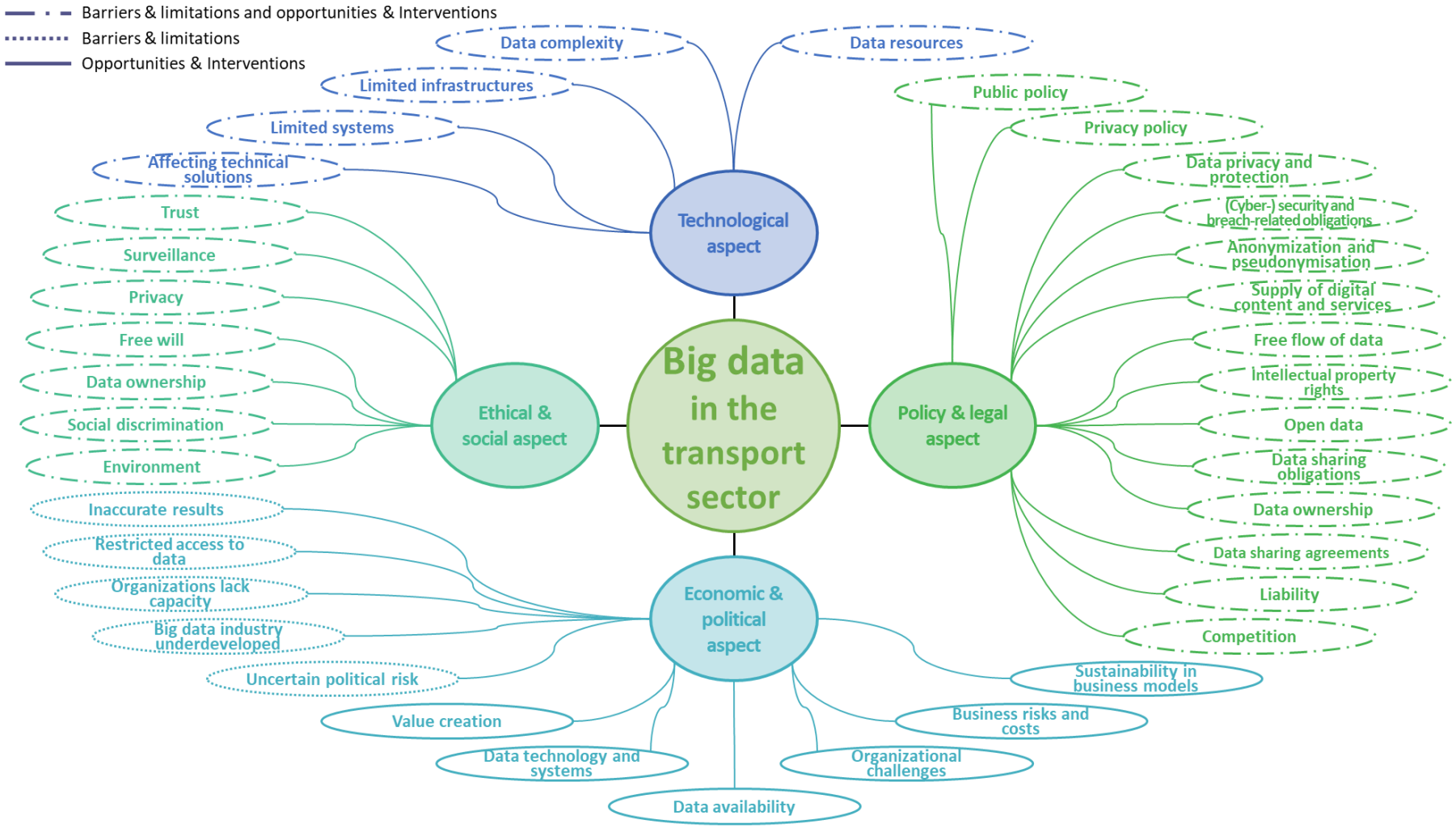


Figure 8 Diagram for barriers/limitations and opportunities/interventions for each aspect

7.3.2 Term distribution based on word count and cloud

This section investigates results by word count and word cloud as shown in the Figure 9, Figure 10 and Figure 11. We considered Bi- and Tri-gram words (e.g., personal data, big data analytics) as a term. As is visible from the figures below, All words or top ranked words appear to be similar for both B/L and O/I.

Figure 9 Word cloud for barriers and limitations

Figure 10 Word cloud for opportunities and interventions

For top 40 words used for word clouds of B/L and O/I, see Appendix B.

Figure 11 Word cloud of top 100 terms from both issues (i.e., BL and OI)

7.3.3 Quantitative analysis via NLP techniques

This section explains how to apply NLP techniques to our SIA in LeMO. For the quantitative analysis, we apply a scientific method by fulfilling the following tasks:

1. Pre-process the descriptions of each item (i.e., barriers/limitations and opportunities/ interventions).
2. Find optimal numbers of topics for the whole text including the descriptions based on minimum topic coherences.
3. Identify text (i.e., the description) belonging to each topic.
4. Count the number of common topics for B/L and O/I in terms of the aspects and issues.
5. Make adjacency matrices for common topics and relation graphs based on the matrices.
6. Show the results through some visualization techniques for easy understanding.

In Step 4, we count four kinds of common topics as follows.

a) Common topics between B/L,
b) Common topics between O/I,
c) Common topics between both B/L and O/I,
d) Common topics between separated both B/L and O/I.

The “c” is counting common topics included in the same titles of B/L and O/I (for example, 'data resources' is included into both B/L and O/I). For the `d`, B/L and O/I, which have the same name, is considered separately.

7.3.3.1 Optimal number of topics

As we discussed in the above Section 7.3, we select an optimal topic number ‘k’ that marks the end of a rapid growth of coherence to find meaningful and interpretable topics. It is called "Elbow method". Following three graphs show the coherences according to topic numbers for B/L, O/I and both contents, respectively.

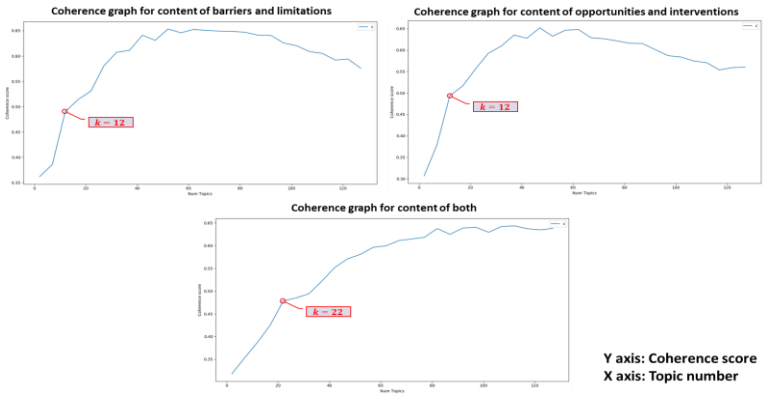


Figure 12 Coherence graph to find optimal topic numbers

We picked a first point which has a higher increase ratio to reduce computation costs and to find more shared factors. That is, the raise ratio of a next point is lower than that of the first point. Detail information could be found in the Appendix B. As a result, dominant topics are assigned for each item and are listed by Table 41, Table 42 and Table 43 in Appendix B.

7.3.3.2 Interaction analysis between issues and aspects

This section analyses the relations (common topics) between issues and aspects. Since quantitative analysis has been used, the numbers of common topics for each issue are simply counted to compose adjacency matrices for B/L, O/I and both. For the 'both' category of B/L and O/I, we have distinguished into two ways. Firstly, items are combined by the issue title (i.e., counting way is 'c' as aforementioned). Secondly, we divided items into each category of B/L or O/L separately like the 'd' way in the above section. The former's abbreviation is BO, and the latter category is called Separated Both (SBO). To show the relations between aspects and issues, we use colourmap[26] and colorbar[27] of Matplotlib. The higher numbers of common topics, the stronger relations and denser blue colour in our maps and bars. The colorbar is used to represent colours and their corresponding topic numbers. The following Figure 13 shows the colour maps and bars for aspect in terms of each category (i.e., BL, OI, BO and SBO).

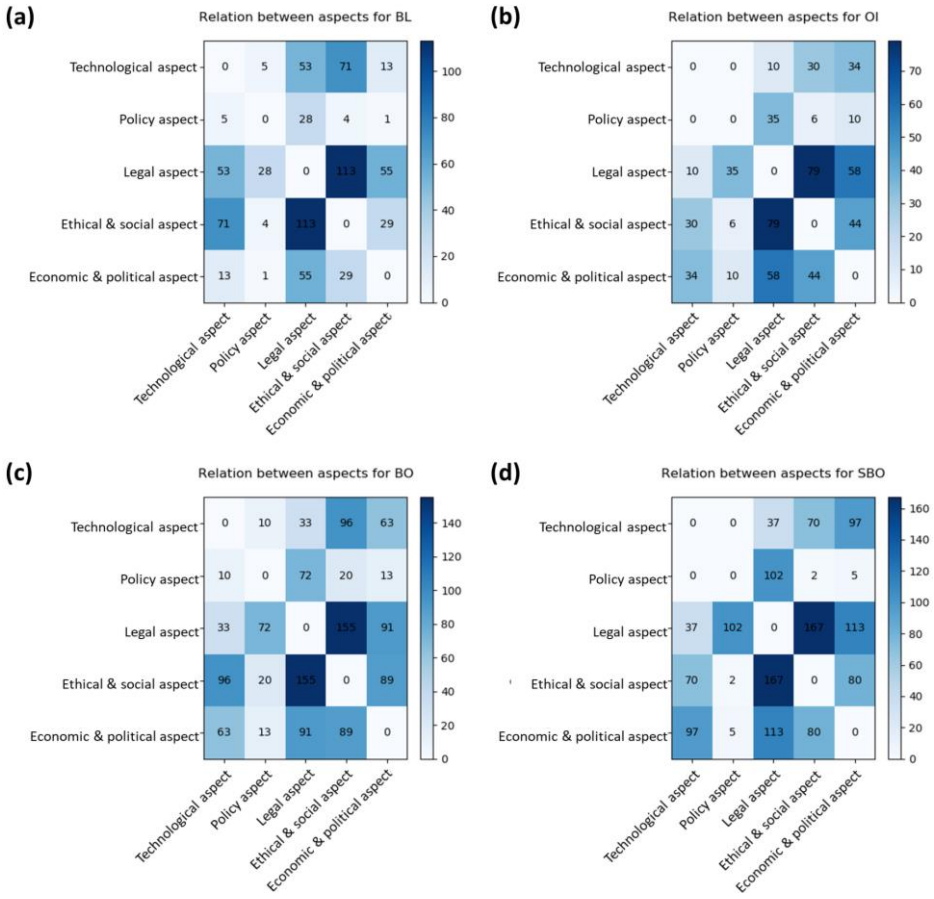


Figure 13 Relation colormaps between aspects

The following colourmaps represent the interactions between issues. Corresponding data for the following Figure 14, Figure 15, Figure 16 and Figure 17 is attached in the Appendix B. In the figures below, the grid divides aspects.

[26] https://matplotlib.org/tutorials/colors/colormaps.html

[27] https://matplotlib.org/3.1.1/tutorials/colors/colorbar_only.html

Figure 14 Relations between issues of barriers and limitations

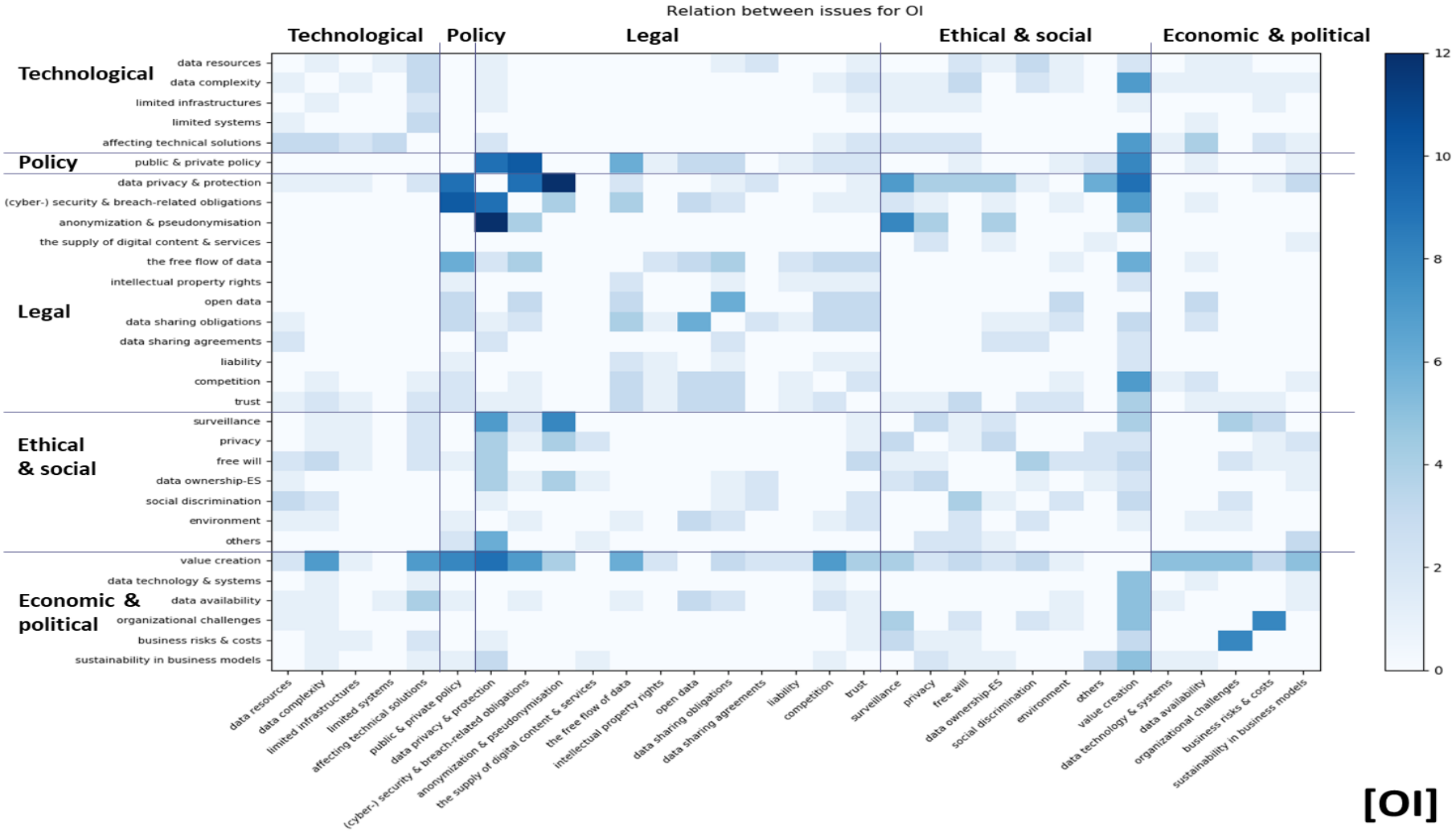


Figure 15 Relations between issues of opportunities and interventions

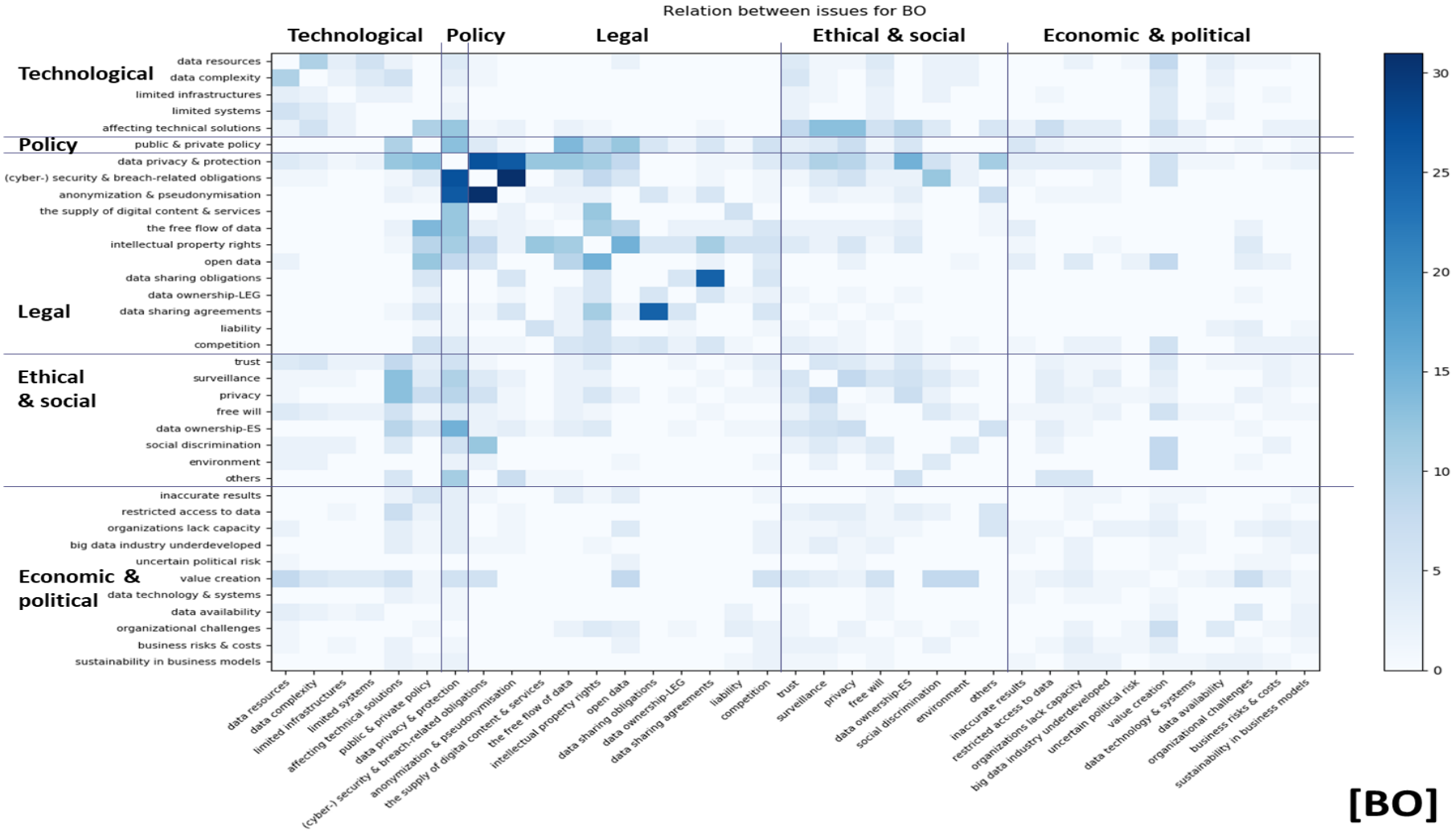


Figure 16 Relations between issues of both (BL and OI)

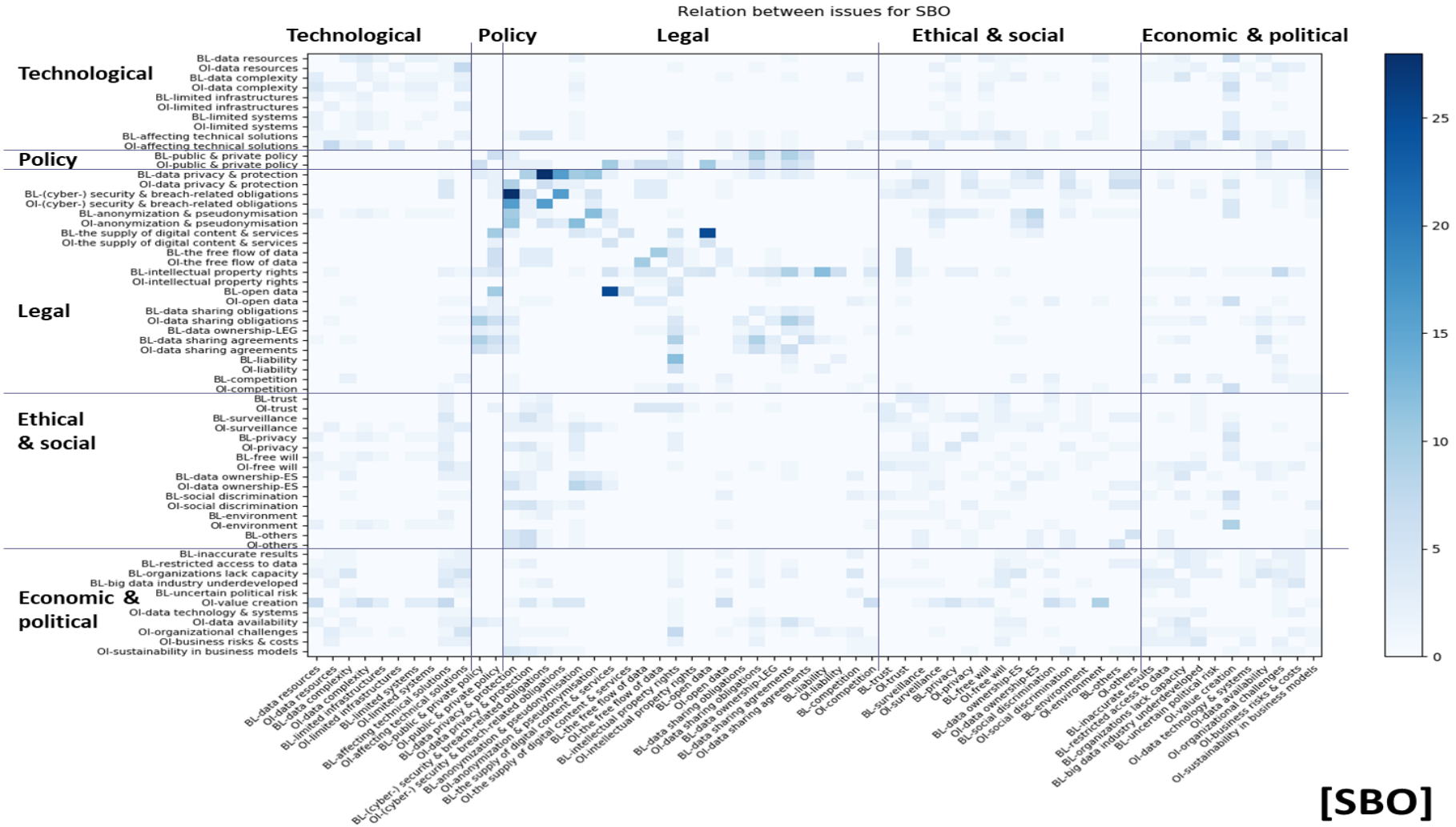


Figure 17 Relations between issues of separated both (BL and OI)

7.3.3.3 Visualization techniques for results' analysis

Moreover, the results applying interactive visualization techniques into our analysis will be updated also in the website as shown in the figures below. The following example shows the results of pyLDAvis [34] for LeMO SIA. This visualization technique provides many useful web interfaces to allow users to find meaningful information about the topic modelling results. For example, the left graph shows distance between topics and the right bar chart lists top-30 terms based on the measure 'Salient'. If a user moves their mouse cursor on a term in the right graph, the left map dynamically shows topics including the corresponding term and represents the term's distribution information among the topics via circle size.

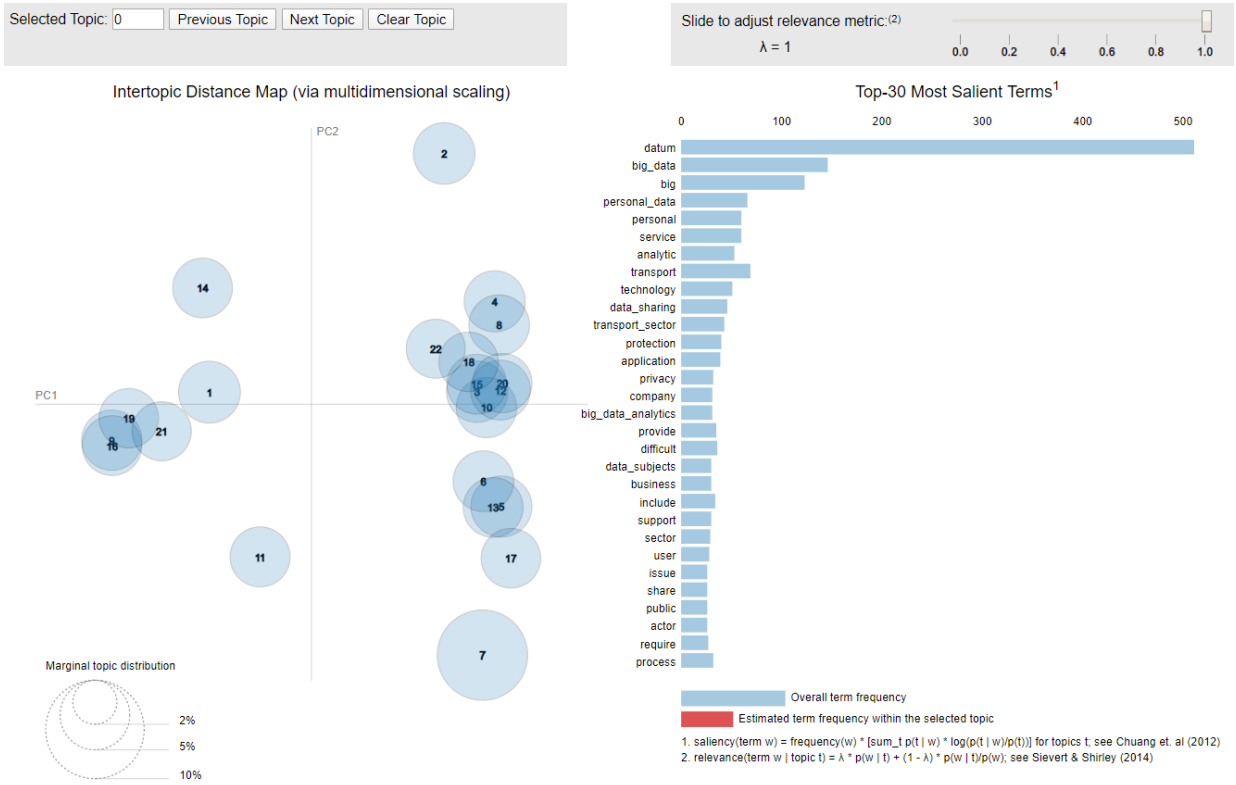


Figure 18 LDAvis visualization example for the topic modelling results of LeMO SIA

Figure 19 and Figure 20 show the force-graph visualisation for the results of common relations between aspects and issues, respectively. Node and edge of this graph indicate items (i.e., issues or aspects) and the number of interactions between the items. Note that the figures' colour was modified to save ink. This visualisation technique also offers interactive interfaces to users. For example, if a user moves a mouse cursor on a node, the list of related issues is pops up. For edges, the number of relations between two nodes is shown.

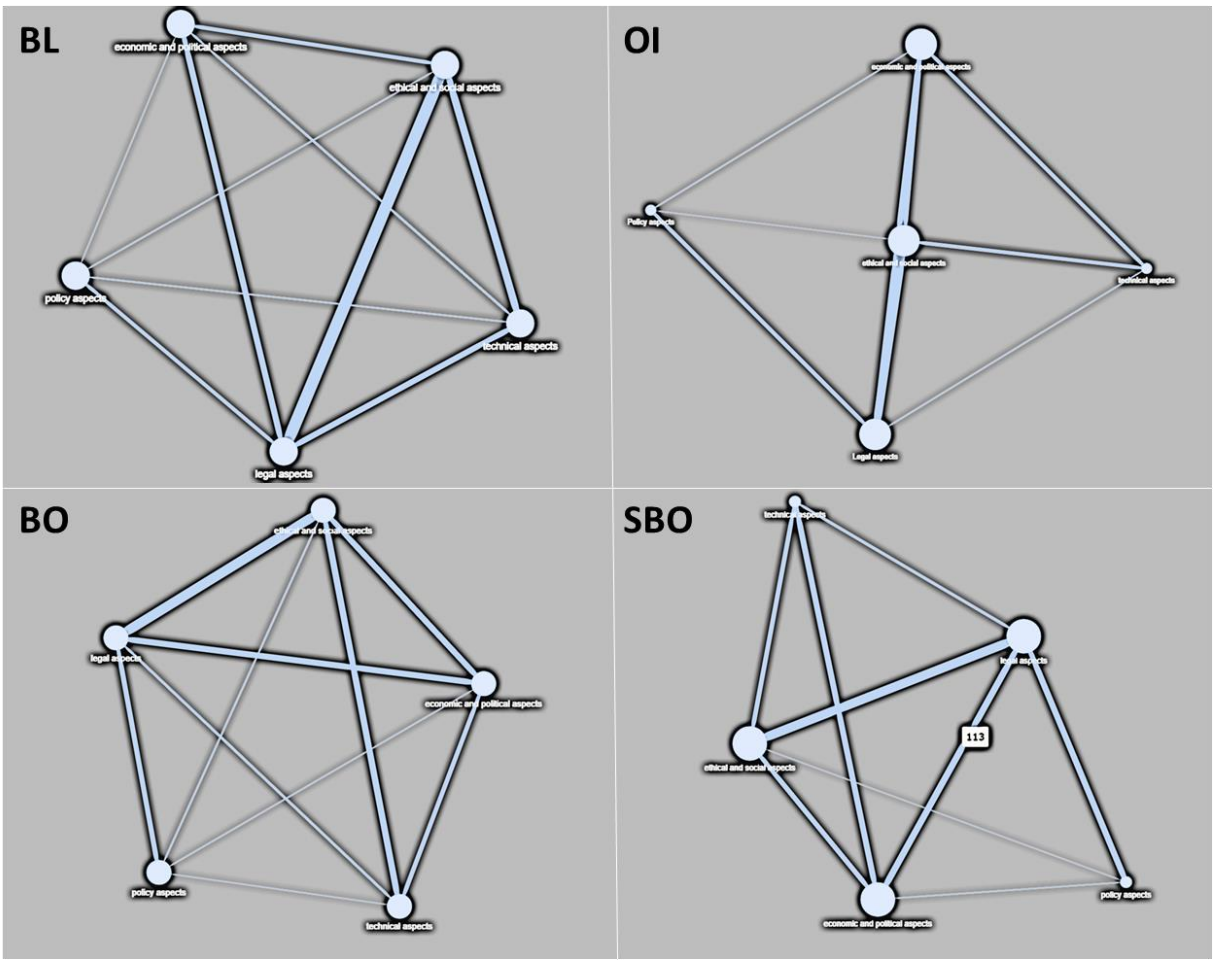


Figure 19 Force-graph for common relation between aspects

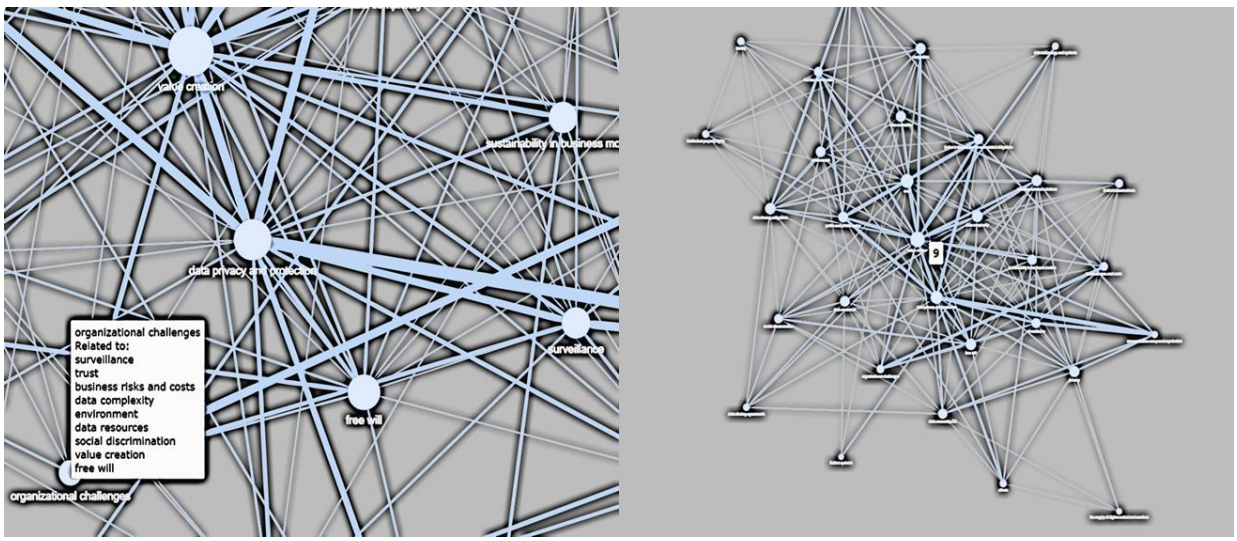


Figure 20 Force-graph for common relation between issues

7.3.3.4 Hypothesis on the analysis of severity and intensity

In this section, we analyse severity of B/L and intensity of O/I based on the following hypothesises about the results of quantitative and qualitative analysis:

- For scores of severity and intensity, we consider qualitative and quantitative analysis results. We use Likert scale [35] for the score (i.e., from 1 (low) to 5 (high)) since it is one of the most used scales.
- In terms of interactions between issues and aspects, it is assumed that the greater the links with others, the greater the severity or intensity.
- Regarding qualitative analysis, if there are relations between items investigated by experts, the relations are considered as more significant than the quantitative results.

Regarding the definitions of the Oxford dictionary, we define the severity as "the fact or condition of barriers/limitations being extremely bad or serious".

Whereas, the intensity indicates the state or quality of being intense for opportunities/ interventions.

7.3.3.5 Data specifications and methodology for the analysis

For this section, we use SBO (i.e., both separated B/L and O/I) for the quantitative analysis. For qualitative analysis, we consider not only the numbers of relations stated in tables for each issue but also the results of an expert discussion.

To fairly consolidate the numbers of interactions based on co-topics and qualitative analysis, we conducted the following processes.

1. First, to avoid wrong interpretation by the impact of self-interactions in qualitative analysis, we eliminated the interactions for each issue.
2. Furthermore, the different item numbers of each issue should be managed to make valances between them. Therefore, interactions of each issue are normalized by division by the number of each issue's items.
3. And then qualitative interactions, which were found by experts, are counted by multiplying by 2.
4. Lastly, we sum the quantitative and qualitative interaction numbers for each issue.

7.3.3.6 Rescaling merged numbers of interactions into severity and intensity

We used a simple statistic formula to re-scale the merged relations into the Likert scale (i.e. from 1 to 5). Given a value scope of rescaling, a maximum observed value O_i^{Max} and a minimum perceived value O_i^{Min} of a i^{th} issue, the rescaled value $\hat{O}_i$ is defined as follows:

$$C = \frac{S^{Max} - S^{Min}}{O_i^{Max} - O_i^{Min}},$$
$$D = S^{Min} - C \times O_i^{Min},$$
$$\hat{O}_i = C \times O_i + D,$$

where S^{Max} and S^{Min} indicate maximum and minimum value of rescaling (i.e. 5 and 1), O_i denotes the observed value for rescaling. Therefore, an issue has 1 as rescaled value if an issue

has a minimum interaction and value 5 is assigned to an issue that has a maximum interaction. It is reasonable because computing interactions considers all issues investigated in our research.

7.3.3.7 Rescaled severity and intensity

For more fruitful discussions, we analysed the situation from three different perspectives, such as interactions with other issues, interactions with others in the same aspect, and interactions with others in the other aspects. Table 35 and Table 36 show the rescaled severity of B/L and intensity of O/I. Computations from mid-processes to results are attached in Appendix B.

Table 35 Rescaled severity for the issues of barrier and limitation

Issue \ Basis	Interactions with others (1)	Interactions with others in the same aspects (2)	Interactions with others in the other aspects (3)
BL-data resources	3.44	3.21	2.81
BL-data complexity	3.17	2.67	3.10
BL-limited infrastructures	3.53	3.38	2.71
BL-limited systems	3.53	3.38	2.71
BL-affecting technical solutions	2.34	1.54	3.41
BL-public and private policy	4.15	1.86	5.00
BL-data privacy and protection	2.36	3.29	1.57
BL-(cyber-) security and breach-related obligations	2.05	2.93	1.63
BL-anonymization and pseudonymisation	2.11	2.19	2.47
BL-the supply of digital content and services	1.83	2.94	1.38
BL-the free flow of data	1.78	2.85	1.43
BL-intellectual property rights	1.27	2.27	1.49
BL-open data	1.65	2.85	1.29
BL-data sharing obligations	3.79	3.92	2.43
BL-data ownership-LEG	2.72	3.52	1.71
BL-data sharing agreements	2.78	2.90	2.43
BL-liability	1.92	3.38	1.00
BL-competition	1.56	1.50	2.62
BL-trust	1.38	2.58	1.29
BL-surveillance	1.60	2.36	1.74
BL-privacy	2.10	2.04	2.62
BL-free will	2.99	2.31	3.29
BL-data ownership-ES	2.25	1.70	3.14
BL-social discrimination	1.74	1.77	2.52
BL-environment	2.99	3.12	2.43
BL-others	2.19	2.58	2.14
BL-inaccurate results	5.00	5.00	2.57
BL-restricted access to data	3.17	2.94	2.81
BL-organizations lack capacity	3.42	3.33	2.66

BL-big data industry underdeveloped	3.35	2.94	3.00
BL-uncertain political risk	3.79	3.65	2.71

Table 36 Rescaled Intensity of the issues of opportunities and interventions

Issue \ Basis	Interactions with others (1)	Interactions with others in the same aspects (2)	Interactions with others in the other aspects (3)
OI-data resources	2.28	2.58	2.24
OI-data complexity	3.17	3.47	2.24
OI-limited infrastructures	2.45	3.65	1.29
OI-limited systems	3.53	3.38	2.71
OI-affecting technical solutions	2.10	2.22	2.43
OI-public and private policy	2.81	1.41	4.05
OI-data privacy and protection	2.21	2.47	2.29
OI-(cyber-) security and breach-related obligations	1.97	3.28	1.17
OI-anonymization and pseudonymisation	3.12	3.92	1.71
OI-the supply of digital content and services	2.72	3.92	1.29
OI-the free flow of data	1.65	2.71	1.43
OI-intellectual property rights	2.19	3.12	1.57
OI-open data	3.53	2.44	3.71
OI-data sharing obligations	3.33	2.85	3.07
OI-data sharing agreements	3.53	3.65	2.43
OI-liability	2.45	3.92	1.00
OI-competition	1.47	2.13	1.86
OI-trust	1.99	1.97	2.57
OI-surveillance	2.12	1.90	2.79
OI-privacy	1.58	2.04	2.07
OI-free will	2.19	1.90	2.86
OI-data ownership-ES	2.54	2.13	3.00
OI-social discrimination	1.74	1.77	2.52
OI-environment	2.72	1.63	3.71
OI-others	2.19	2.58	2.14
OI-value creation	1.00	1.00	2.55
OI-data technology and systems	4.60	4.73	2.43
OI-data availability	4.51	3.47	3.67
OI-organizational challenges	2.72	2.58	2.71
OI-business risks and costs	2.81	3.21	2.14
OI-sustainability in business models	2.19	2.31	2.43

7.4 Analysis and generic observations from a comparative study of tables in barriers/limitations and opportunities/interventions

7.4.1 Technological aspect

Within technological aspect a perusal through the barriers/limitation on the one hand and the opportunities and intervention, on the other hand, reveals the following:

1) Though many barriers/limitations fall in "severe" category, the corresponding interventions and opportunities suggested in the present deliverable can help in overcoming most of the barriers;
2) Some barriers (such as fragmented data sources, choosing the right infrastructure, etc.) can be overcome by relying on existing technological/research solutions (such as by using data platforms that consolidate data and by referring to research on benchmarked systems);
3) While other barriers (such as limitations of existing traffic management systems and lack of data quality) would have to be overcome through the development of new frameworks, workflows, and structures;
4) It also emerges that some of the barriers (such as limitations of infrastructure capabilities and high costs) can be overcome through the use of cloud-based services as well as open source technology. However, it would have to be seen if this solution will offer adequate security/safety. Similarly, the lack of skilled workers/experts can be addressed by relying on emerging working models that base on freelancers and remote workers. These experts can help build up the big data solution with their expertise and transfer it on to full-time employees during this time.
5) Lastly, specific research in some areas, as well as the establishment of education programmes and training opportunities, can help foster more effective use of big data in the transport sector and beyond.

7.4.2 Policy and legal aspect

On policy and legal aspect, the following observations can be made:

1) On the policy level though there is an apparent conflict between sharing data and protecting privacy, a general tendency can be discerned. Confidence is being instilled among people to use more technology driven solutions, by reinforcing privacy. This can also help overcome the trust related barrier under ethical & social aspect.
2) Openness of non-personal data is also being promoted at the policy level and governments are being nudged to share their data to create a level playing field. Similarly, soft laws that promote sharing of data within private sector have also been observed. This is helpful for overcoming the economic and political barrier related to data silos.
3) On the legal side, it emerges that due to the applicability of different regulations, customs and practices there are numerous barriers/limitations related to the legal aspect and relatively less opportunities/interventions. This could be an indicative of a very complex and ambiguous regulatory framework, which does not offer many

opportunities and may need strong interventions to facilitate the use of big data in the transport sector.

4) The stringencies and complexities introduced by regulations such as the GDPR, Free-flow regulation, PSI Directive and NIS Directive would need to be overcome through different steps, including:
   a. Developing coordinated and EU-wide guidance templates, that take into account complex data processing activities envisaged by big data technologies.
   b. Using new technologies that balance competing concerns and allow easy understanding and exercise of individual rights.
   c. Different stakeholders (authorities, standardization bodies, service providers, vendors and industry players) coming together to develop standards, certification mechanisms, seals, marks and codes of conduct that improve compliance with various legislations.
   d. Development of better anonymization techniques to help comply with regulations and to facilitate data sharing between stakeholders in the big data value chain.
   e. Adopting legislation, developing better Intellectual Property Regulations; ensuring interoperability through technical specifications and standards; and using public tendering to prompt data sharing among stakeholders of the big data value chain.
   f. Adopting clear regulations on (extra-contractual and statutory) liability to reinforce legal certainty.

7.4.3 Ethical and social aspect

On ethical and social aspect following general observations can be made:

1) Lack of clarity on data ownership is a concern. The existing data-related legal rights do not respond sufficiently to the needs of the actors in the data value cycle, multiple actors in the data value cycle stake their claim to own the data. Use of specific techniques such as blockchain could offer a possible intervention by tracking clear ownerships and can help declutter the situation.
2) 'Lack of data quality' has been mentioned as a barrier/limitation under technological as well as economic and political aspect and it is considered as barriers/limitations related to social discrimination and trust under the ethical and social aspect. The intervention offered under the technological aspect, which suggests the creation of data curation frameworks and workflows to improve the quality of data, could help overcome these barriers/limitations. Also, suggestions under ethical and social aspect on improving and augmenting data sources could be useful. Similarly, use of public data, as well as its validation via data aggregation and cross-analysis, could improve the low quality of data.
3) Threat to free will, privacy and data ownership related concerns emerge as major barriers/limitations under ethical and social aspect. However, several opportunities/ interventions (such as the use of threshold systems, encryption and anonymization techniques as well as effective implementation of regulations) could alleviate negative influences of these barriers/limitations. In addition, benefits in relation to safety,

efficiency and personalisation of services could allow societies to strike a compromise and balance in situations where competitive concerns emerge.

4) Regarding concerns and misunderstanding about exploiting big data technologies, there could be several interventions having a generally positive effect and increasing trust within the utilisation of big data technologies. For example, providing transparent information to individuals whose data is used; educating the general public and removing misconceptions on personal data protection and providing greater control to data subjects through close involvement.
5) Overall, the opportunities/interventions related to the use of big data in the transport sector exceed the barriers/limitations under the ethical and social aspect. The use of big data in the transport sector could lead to more personalised and efficient services, could help addressing environmental concerns and reduce social discrimination.

7.4.4 Economic and political aspect

For economic and political aspect, the following broad observations can be made:

1) The value creation opportunities afforded by big data, both in the use of big data and in the sale/supply of big data, motivates the industry and public sector in pursuing big data technology.
2) However, there are broadly speaking 5 mechanisms which negative impact realisation of these opportunities throughout the transport sector. These include 'Inaccurate results', 'Restricted access to data', 'Lacking institutional capacity', 'Underdeveloped big data industry' and 'Uncertain political risk'.
3) These mechanisms can translate into risks and negative economic impacts for businesses including poor results, bad products and services, loss of reputation and customers, risks associated to security breaches. They also impact the competitiveness of the European big data industry against other countries. And they can lead to technocratic and algorithmic governance of transport, putting vulnerable and marginal groups at risks.
4) Many organizations aiming to implement big data technology need assistance directly, as well as indirectly through research and policies that entire sectors can benefit from.
5) Similarly, interventions listed under the economic & political aspect, especially those related to data availability (OPT-EP-DA), organizational challenges (OPT-EP-OC), business risk and cost (OPT-EP-BRC), sustainability in business models (OPT-EP-SBM) and data technology & systems (OPT-EP-DTS), could also be used to address the limitations and barriers under technical, policy & legal, and ethical & social aspects.

8 Conclusion

In this deliverable we discussed the opportunities/interventions related to the four different aspects listed below:

- Technological;
- Policy and legal;
- Ethical and social; and
- Economic and political

Some of the opportunities/interventions listed here compliment the barriers/limitations for each aspect listed in D.4.1. The deliverable also provides an impact analysis of the various barriers/limitations and opportunities/interventions looking, comparing their volume within and across different aspects.

NLP techniques have been used to conduct this analysis based on the total of 243 items across 62 issues as barriers/limitations and opportunities/interventions of technological, policy & legal, ethical & social and economic & political aspects on the use of big data in the transport sector. A comparative reading of the tables and discussions complement the results from applying NLP techniques. As a result, Section 7.6 as a summary concluded with the overall results of socio-economic impact analysis.

The horizontal analysis illustrates that big data does not stand for the same practice in every case study domain but covers a wide variety of datasets and data practices and the technical challenges these present. Mapped across the big data value chain the technological/policy and legal/social and ethical/economic and political challenges raised in the case studies are mostly observed in the data collection, data curation but in a lesser extent in the data usage phase. Often these challenges are not simply challenging but complexing in terms of societal views. The overall picture shows positive benefits but also the potential to negatively affect other important values in the other aspects.

The outcomes of this analysis will form the basis for the next tasks in WP 4. The next task of the work package will build upon the analysis presented in this deliverable and use the suggested interventions to construct a research and policy roadmap for facilitating the use of big data in the transport sector.

References

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

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

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

[4] Big Data Value Association, Ed., 'European Big Data Value - Strategic Research and Innovation Agenda', Oct-2017.

[5] 'ISO/IEC TR 20547-2:2018 - Information technology - Big data reference architecture - Part 2: Use cases and derived requirements'.

[6] 'ISO/IEC JTC 1/SC 42 - Artificial intelligence', ISO. [Online]. Available: http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/committee/67/94/6794475.html. [Accessed: 26-Nov-2019].

[7] K. Hee et al., 'Understanding and Mapping Big Data in Transport Sector', 2018.

[8] C. Katrakazas, 'NOESIS - D2.1 Big Data and emerging transportation challenges', 30-Jun-2018.

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

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

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

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

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

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

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

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

[17] IBM, 'IBM SPSS Modeler CRISP-DM Guide', 2011.

[18] K. Hee, T. Ivanov, R. V. Zicari, and R. Waldenfels, 'Big Data Methodologies, Tools and Infrastructures', 2018. [Online]. Available: https://static1.squarespace.com/static/59f9cdc2692ebebde4c43010/t/5b6d4674032b

e489a442fa8b/1533888127770/20180716_D1.3_Big+data+methodologies%2C+tools+and+infrastructures_LeMO.pdf. [Accessed: 05-Aug-2019].

[19] M. Hong, R. Akerkar, T. Teoh, T. Eichhorn, and J. Debussche, 'Report on the characterization of the barriers and limitations', 30-September-2019. [Online]. Available: https://lemo-h2020.eu/s/20190930_D41_Report-on-the-characterization-of-the-barriers-and-limitation-final.pdf

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

[21] J. L. Berral, N. Poggi, D. Carrera, A. Call, R. Reinauer, and D. Green, 'ALOJA: A Framework for Benchmarking and Predictive Analytics in Big Data Deployments', IEEE Trans. Emerg. Top. Comput., vol. 5, no. 4, pp. 480–493, Oct. 2017.

[22] B. Pernici et al., 'Relating Big Data Business and Technical Performance Indicators', in Conference of the Italian Chapter of AIS, 2018, pp. 1–12.

[23] E. Ciceri, M. Italia, M. Mosconi, M. Italia, M. Önen, and O. Ermis, 'PAPAYA: A Platform for Privacy Preserving Data Analytics', p. 2, Jul. 2019.

[24] T. Loruenser, D. Slamanig, T. Länger, and H. C. Pöhls, 'PRISMACLOUD Tools: A Cryptographic Toolbox for Increasing Security in Cloud Services', in 2016 11th International Conference on Availability, Reliability and Security (ARES), 2016, pp. 733–741.

[25] P. Gouvas and A. Michalas, 'PaaSword Reference Architecture - Deliverable D1.3', Sep-2015.

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

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

[28] Lucassen, B. Klievink, and L. Tavasszy, "A Living Lab Framework : facilitating the adoption of innovations in international information infrastructures," in Transport Research Arena 2014, 2014.

[29] V. Ramaswamy and F. Gouillart, "Building the Co- Creative Enterprise," Harv. Bus. Rev., no. October 2010, 2010.

[30] P. Sarkar, "Introduction to DaaS," in Data as a Service:A Framework for Providing Reusable Enterprise Data Services, 2015.

[31] B. Van Wee and S. Roeser, "Ethical Theories and the Cost–Benefit Analysis-Based Ex Ante Evaluation of Transport Policies and Plans," Transp. Rev., vol. 33, no. 6, pp. 743–760, Nov. 2013.

[32] J. Höchtl, P. Parycek, and R. Schöllhammer, “Big data in the policy cycle: Policy decision making in the digital era,” J. Organ. Comput. Electron. Commer., vol. 26, no. 1–2, pp. 147–169, Apr. 2016.

[33] Mackenzie Valley Environmental Impact Review Board, ‘Socio-Economic Impact Assessment GUIDELINES’, March 2007. [Online] Available: http://reviewboard.ca/file/1024/download?token=1DDLP3jP

[34] Sievert, Carson, and Kenneth Shirley, ‘LDAvis: A method for visualizing and interpreting topics’, In Proceedings of the workshop on interactive language learning, visualization, and interfaces, pp. 63-70. 2014.

[35] Likert, R. ‘A technique for the measurement of attitude scales’, 1932.

[36] R. Akerkar, P.S. Sajja. Intelligent Techniques for Data Science, Springer International Publishing, DOI 10.1007/978-3-319-29206-9, 2016.

Appendix A: List of Reference Notations Considered

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

Table 37 List of reference notations considered

Code	Description
OPT-TM-DR-#	Opportunities related to data resources in technological aspect
OPT-TM-DC-#	Opportunities related to data complexity in technological aspect
OPT-TM-LI-#	Opportunities related to limited infrastructure in technological aspect
OPT-TM-IS-#	Opportunities related to limited infrastructures and systems in technological aspect
OPT-TM-AT-#	Opportunities related to affecting technical solutions in technological aspect
OPT-POL-PU-#	Opportunities related to public policy in policy aspect
OPT-POL-PR-#	Opportunities related to privacy policy in policy aspect
OPT-LEG-DP-#	Opportunities related to data privacy and protection in legal aspect
OPT-LEG-CBO-#	Opportunities related to (cyber-) security and breach-related obligations in legal aspect
OPT-LEG-AP-#	Opportunities related to anonymization and pseudonymisation in legal aspect
OPT-LEG-DIG-#	Opportunities related to the supply of digital content and services in legal aspect
OPT-LEG-FF-#	Opportunities related to the free flow of data in legal aspect
OPT-LEG-IP-#	Opportunities related to intellectual property rights in legal aspect
OPT-LEG-OD-#	Opportunities related to open data in legal aspect
OPT-LEG-DS-#	Opportunities related to data sharing obligations on legal aspect
OPT-LEG-DO-#	Opportunities related to data ownership in legal aspect
OPT-LEG-DSA-#	Opportunities related to data sharing agreements in legal aspect
OPT-LEG-LI-#	Opportunities related to liability in legal aspect
OPT-LEG-COM-#	Opportunities related to competition in legal aspect
OPT-ES-TR-#	Opportunities related to trust in ethical and social aspect
OPT-ES-SUR-#	Opportunities related to surveillance in ethical and social aspect
OPT-ES-PRI-#	Opportunities related to privacy in ethical and social aspect
OPT-ES-FW-#	Opportunities related to free will in ethical and social aspect
OPT-ES-DO-#	Opportunities related to data ownership in ethical and social aspect
OPT-ES-SD-#	Opportunities related to social discrimination in ethical and social aspect
OPT-ES-ENV-#	Opportunities related to environment in ethical and social aspect
OPT-ES-OT-#	Opportunities related to others in ethical and social aspect
OPT-EP-VC-#	Opportunities related to value creation in economic and political aspect
OPT-EP-DTS-#	Opportunities related to data technology and systems in economic and political aspect
OPT-EP-DA-#	Opportunities related to data availability in economic and political aspect
OPT-EP-OC-#	Opportunities related to organizational challenges in economic and political aspect
OPT-EP-BRC-#	Opportunities related to business risk and cost in economic and political aspect
OPT-EP-SBM-#	Opportunities related to sustainability in business models in economic and political aspect

Appendix B: Detailed Results of Applying NLP Techniques for SIA

To understand these word clouds, we used top 40 words for Barriers/Limitations (B/L) and Opportunities/Interventions (O/I) as shown in the below table.

Table 38 Top 40 words for B/L and O/I

Rank	Word (B/L)	Word (O/I)	Rank	Word (B/L)	Word (O/I)
1	*Data*	*Data*	**21**	Issue	*Data subject*
2	*Big data*	*Big data*	**22**	Service	**Open data**
3	Personal data	Service	**23**	Actor	**Value**
4	Protection	Transport	**24**	**Copyright**	**Infrastructure**
5	*Analytic*	*Analytic*	**25**	**Content**	**System**
6	Application	Personal data	**26**	**Database**	**Level**
7	*Technology*	*Technology*	**27**	**Dataset**	**Management**
8	Transport	Data sharing	**28**	**Risk**	Protection
9	Data sharing	Transport sector	**29**	**Analysis**	Actor
10	*Big data analytics*	*Big data analytics*	**30**	**Big data context**	**Cost**
11	**Context**	Information	**31**	Business	**Customer**
12	**User**	**Company**	**32**	**Case**	**Framework**
13	**Directive**	Business	**33**	**Ownership**	Issue
14	Organisation	**Opportunity**	**34**	Sector	**Model**
15	**Right**	**Case study**	**35**	**Security**	**Network**
16	Transport sector	**Industry**	**36**	**Digital content**	Organisation
17	**EU**	Privacy	**37**	**Intellectual property**	**Requirement**
18	Information	***Research***	**38**	**Lack**	**Solution**
19	Privacy	Sector	**39**	**Legal uncertainty**	**Time**
20	*Data subject*	Application	**40**	**Legislation**	**Data collection**

In the above table, terms under B/L and O/I that are same and have similar ranks are remarked by *italic* font. Terms that are same and don't have similar rank are marked by Normal font. Lastly, the following figure is a word cloud including top 100 terms from both B/L and O/I. For more details, see Table 39 in Appendix B.

The below table shows the word and its number for Figure 9, Figure 10 and Figure 11. In this appendix, we use BL, OI, and BO abbreviation for Barriers/Limitations, Opportunities/Interventions and Both (i.e., Barriers/Limitations and Opportunities/ Interventions). ***Note that all following data might be incomplete or not be word since we used the same results from applying NLP techniques to avoid any distortion.***

Table 39 Detail results for word cloud

BL		OI		BO	
Co.	word	Co.	word	Co.	word
301	data	209	data	510	data
101	big_data	57	big_data	151	big_data
40	personal_data	44	service	62	personal_data
26	protection	38	transport	61	transport
25	analytic	24	analytic	58	service
23	application	24	personal_data	49	analytic
23	technology	24	technology	47	technology
23	transport	22	data_sharing	42	data_sharing

20	data_sharing	22	transport_sector	38	big_data_analytics
19	big_data_analytics	21	big_data_analytics	38	protection
19	context	20	information	38	transport_sector
17	user	19	company	37	application
16	directive	16	business	35	information
16	organisation	16	opportunity	30	privacy
16	right	15	case_study	28	company
16	transport_sector	15	industry	28	data_subject
15	EU	15	privacy	27	business
15	information	15	research	27	organisation
15	privacy	15	sector	26	sector
14	data_subject	14	application	26	user
14	issue	14	data_subject	25	EU
14	service	14	open_data	25	issue
13	actor	14	value	24	actor
13	copyright	13	infrastructure	24	case_study
12	content	13	system	24	value
12	database	12	level	21	context
12	dataset	12	management	21	right
12	risk	12	protection	21	system
11	analysis	11	actor	20	requirement
11	big_data_context	11	cost	20	time
11	business	11	customer	19	directive
11	case	11	framework	19	infrastructure
11	ownership	11	issue	19	level
11	sector	11	model	19	model
11	security	11	network	19	subject
10	digital_content	11	organisation	18	industry
10	intellectual_property	11	requirement	18	risk
10	lack	11	solution	18	security
10	legal_uncertainty	11	time	17	analysis
10	legislation	10	data_collection	17	obligation
10	non_personal	10	data_protection	17	opportunity
10	obligation	10	data_value	17	ownership
10	value	10	EU	16	activity
9	anonymisation	10	GDPR	16	business_model
9	company	10	platform	16	collection
9	discrepancy	10	real_time	16	data_sources
9	property	10	subject	16	quality
9	requirement	9	activity	16	real_time
9	subject	9	assessment	16	research
9	term	9	collection	16	source
9	time	9	quality	16	value_chain
8	access	9	user	15	access
8	algorithm	9	vehicle	15	assessment
8	anonymisation_techniques	8	chain	15	case
8	copyright_protection	8	decision	15	content
8	model	8	development	15	cost
8	party	8	free_flow	15	GDPR
8	practice	8	individual	15	management
8	protection_legislation	8	policy	15	solution
8	PIS_directive	8	policy_framework	15	term
8	source	8	provider	14	anonymisation_techniques
8	system	8	public_transport	14	customer
7	account	8	service_providers	14	framework
7	activity	8	source	14	non_personal

7	agreement	8	traffic	14	traffic
7	collection	7	access	13	copyright
7	investment	7	obligation	13	database
7	level	7	security	13	dataset
7	purpose	7	trust	13	investment
7	quality	6	analysis	13	legislation
7	state	6	anonymisation_techniques	13	policy
7	uncertainty	6	change	12	anonymisation
6	assessment	6	cloud	12	development
6	authority	6	control	12	digital_content
6	business_model	6	data_analytics	12	free_flow
6	competent_authorities	6	data_platforms	12	NIS_directive
6	end_users	6	driven_applications	12	party
6	infrastructure	6	increase	12	vehicle
6	legal_instruments	6	investment	11	big_data_context
6	member	6	legal_certainty	11	chain
6	NIS_directive	6	machine_learning	11	control
6	real_time	6	need	11	decision
6	scope	6	NIS_directive	11	network
6	situation	6	ownership	11	people
6	third_parties	6	people	11	platform
6	traffic	6	personal_information	11	practice
6	type	6	private_actors	11	technique
5	behaviour	6	risk	11	trust
5	concept	6	start_ups	10	account
5	control	6	technique	10	agreement
5	flow	6	term	10	authority
5	GDPR	6	traffic_management	10	data_analytics
5	measure	6	transport_systems	10	end_users
5	people	5	compliance	10	individual
5	policy	5	impact	10	intellectual_property
5	possibility	5	innovation	10	lack
5	project	5	program	10	legal_uncertainty
5	size	5	right	10	open_data
5	technique	5	standard	10	property
5	worker	5	storage	10	provider
4	amount	5	transparency	10	public_transport
4	approach	4	approach	10	service_providers
4	arrangement	4	authority	10	situation
4	border	4	barrier	10	traffic_management

The following table lists coherence values according to topics numbers for Figure 12.

Table 40 Detail results for coherence graph

	BL		OI		BO	
Topic No.	**Coherence**	**increase**	**Coherence**	**increase**	**Coherence**	**increase**
2	0.3617		0.3067		0.3179	
7	0.3859	0.0242	0.3793	0.0726	0.3532	0.0353
12	0.4901	**0.1042**	0.4938	**0.1145**	0.3874	0.0342
17	0.5140	0.0239	0.5164	0.0226	0.4256	0.0382
22	0.5308	0.0168	0.5560	0.0396	0.4782	**0.0526**
27	0.5805	0.0497	0.5927	0.0367	0.4850	0.0068
32	0.6074	0.0269	0.6092	0.0165	0.4941	0.0091
37	0.6111	0.0037	0.6353	0.0261	0.5217	0.0276

42	0.6409	0.0298	0.6277	-0.0076	0.5516	0.0299
47	0.6310	-0.0099	0.6520	0.0243	0.5709	0.0193
52	0.6533	0.0223	0.6325	-0.0195	0.5807	0.0098
57	0.6457	-0.0076	0.6464	0.0139	0.5963	0.0156
62	0.6526	0.0069	0.6483	0.0019	0.5995	0.0032
67	0.6503	-0.0023	0.6289	-0.0194	0.6112	0.0117
72	0.6491	-0.0012	0.6265	-0.0024	0.6147	0.0035
77	0.6487	-0.0004	0.6217	-0.0048	0.6184	0.0037
82	0.6467	-0.0020	0.6160	-0.0057	0.6375	0.0191
87	0.6411	-0.0056	0.6153	-0.0007	0.6251	-0.0124
92	0.6406	-0.0005	0.6011	-0.0142	0.6387	0.0136
97	0.6256	-0.0150	0.5871	-0.0140	0.6404	0.0017
102	0.6203	-0.0053	0.5839	-0.0032	0.6295	-0.0109
107	0.6087	-0.0116	0.5746	-0.0093	0.6420	0.0125
112	0.6052	-0.0035	0.5708	-0.0038	0.6436	0.0016
117	0.5920	-0.0132	0.5537	-0.0171	0.6373	-0.0063
122	0.5939	0.0019	0.5592	0.0055	0.6347	-0.0026
127	0.5757	-0.0182	0.5603	0.0011	0.6379	0.0032
Max	**0.6533**	**0.1042**	**0.6520**	**0.1145**	**0.6436**	**0.0526**

The following tables list topics for each issue for B/L, O/I and both. The contribution means the percentage proportion of the topic in the given issue. Detailed topic information is listed in the Appendix B. The numbers in the column 'Dominant topics' indicates the topic ID in Table 44. Keywords related to each topic are listed in the table (refer to Appendix B).

Table 41 Dominant topics for barrier & limitation issues

Aspect	Issue	Dominant topics (Contribution)
Technical	data resources	9 (0.2019), 9 (0.1104), 10 (0.1315)
	data complexity	0 (0.1221), 9 (0.1702), 10 (0.1283)
	limited infrastructures	9 (0.1658)
	limited systems	10 (0.1643)
	affecting technical solutions	9 (0.1031), 2 (0.1296), 9 (0.2151), 9 (0.1724), 11 (0.1426), 9 (0.1532), 3 (0.1652)
Policy	public and privacy policy	0 (0.2553), 0 (0.1452), 10 (0.124)
Legal	data privacy and protection	8 (0.1869), 8 (0.2861), 8 (0.1205), 8 (0.2295), 10 (0.1101), 4 (0.1837), 2 (0.1256), 3 (0.1818), 10 (0.1143), 1 (0.1618), 8 (0.1484), 1 (0.1872), 5 (0.1658), 5 (0.161)
	(cyber-) security and breach-related obligations	10 (0.1309), 2 (0.1296), 2 (0.1054), 1 (0.2899), 1 (0.1946), 1 (0.1701), 7 (0.1322), 7 (0.2618), 7 (0.2356), 0 (0.1876)
	anonymization and pseudonymisation	8 (0.1818), 9 (0.1338), 10 (0.162), 6 (0.1791), 8 (0.1691), 2 (0.1194), 8 (0.1406)
	the supply of digital content and services	11 (0.1913), 11 (0.2943), 11 (0.1678), 7 (0.1924), 8 (0.164), 11 (0.2548)
	the free flow of data	8 (0.2334), 0 (0.1669), 0 (0.1667), 0 (0.2193)
	intellectual property rights	5 (0.1888), 5 (0.2272), 7 (0.1675), 5 (0.1906), 2 (0.1304), 5 (0.192), 5 (0.1707), 5 (0.2799), 5 (0.178), 4 (0.1239), 5 (0.1213), 0 (0.1305), 0 (0.3008), 0 (0.3012)
	open data	4 (0.2251), 0 (0.1254), 7 (0.2835), 7 (0.1201), 2 (0.1471)
	data sharing obligations	0 (0.1945)
	data ownership-LEG	6 (0.1157), 0 (0.2336)
	data sharing agreements	10 (0.1455), 7 (0.1418), 10 (0.1223), 7 (0.1698), 10 (0.171)

	liability	4 (0.1703), 7 (0.1341)
	competition	10 (0.1406), 1 (0.1056), 7 (0.1009)
Ethical and social	trust	2 (0.1365), 7 (0.0995), 2 (0.1435)
	surveillance	9 (0.1499), 2 (0.141), 9 (0.1789), 9 (0.1368), 2 (0.1329)
	privacy	4 (0.1155), 11 (0.1374), 9 (0.1397)
	free will	9 (0.1432), 2 (0.1086)
	data ownership-ES	0 (0.1273), 8 (0.1956), 9 (0.1329), 9 (0.0962)
	social discrimination	6 (0.1063), 3 (0.1097), 6 (0.1316)
	environment	2 (0.1338)
	others	8 (0.1615), 10 (0.1723), 10 (0.1818)
Economic and political	inaccurate results	3 (0.1735), 3 (0.17)
	restricted access to data	10 (0.1682), 5 (0.1062), 3 (0.1587)
	organizations lack capacity	6 (0.291), 3 (0.2287), 3 (0.1306), 3 (0.1992), 3 (0.145)
	big data industry underdeveloped	5 (0.1585), 2 (0.1236), 2 (0.1219)
	uncertain political risk	2 (0.1615)

Table 42 Dominant topics for opportunity and intervention issues

Aspect	Issue	Dominant topics (Contribution)
Technological	data resources	7 (0.1588), 5 (0.139), 6 (0.1731)
	data complexity	2 (0.1595), 3 (0.1848), 5 (0.2593)
	limited infrastructures	2 (0.173)
	limited systems	7 (0.1642)
	affecting technical solutions	2 (0.1697), 2 (0.1808), 7 (0.1361), 7 (0.2126), 7 (0.219), 3 (0.2113)
Policy	public and privacy policy	8 (0.1256), 11 (0.1323), 0 (0.1349), 0 (0.2066), 0 (0.2579), 10 (0.1495)
Legal	data privacy and protection	6 (0.1216), 11 (0.2105), 11 (0.1669), 2 (0.1586), 4 (0.1973), 11 (0.1998), 4 (0.2213), 0 (0.1849), 0 (0.3119), 4 (0.1554)
	(cyber-) security and breach-related obligations	0 (0.1294), 10 (0.1189), 0 (0.3384), 4 (0.1321), 0 (0.2344)
	anonymization and pseudonymisation	4 (0.1763), 4 (0.1492), 4 (0.2628), 4 (0.1901)
	the supply of digital content and services	1 (0.2895)
	the free flow of data	0 (0.2533), 8 (0.1297), 8 (0.1521), 10 (0.1492)
	intellectual property rights	8 (0.1571)
	open data	10 (0.199), 10 (0.1695), 10 (0.1564)
	data sharing obligations	10 (0.1449), 10 (0.1103), 8 (0.1685), 6 (0.1741)
	data sharing agreements	6 (0.1319), 6 (0.1614)
	liability	8 (0.1195)
	competition	3 (0.1147), 8 (0.1907), 10 (0.2112)
Ethical and social	trust	5 (0.12), 8 (0.153), 10 (0.1207), 2 (0.1468)
	surveillance	4 (0.2277), 2 (0.1359), 4 (0.1311), 9 (0.1723)
	privacy	1 (0.145), 1 (0.1254), 2 (0.1068), 4 (0.1354)
	free will	5 (0.1185), 5 (0.1302), 2 (0.1122), 11 (0.1607)
	data ownership-ES	6 (0.1161), 4 (0.1504), 1 (0.143)
	social discrimination	6 (0.1053), 5 (0.1293), 5 (0.1134)
	environment	5 (0.1288), 10 (0.1544)

	others	11 (0.2005), 1 (0.1325), 11 (0.1271)
Economic and political	value creation	8 (0.1138), 4 (0.1301), 3 (0.2737), 5 (0.1239), 0 (0.1272), 3 (0.1218), 3 (0.1995), 2 (0.1805), 8 (0.1434), 3 (0.149), 3 (0.1204), 0 (0.167), 6 (0.2053), 9 (0.1432)
	data technology and systems	3 (0.1388)
	data availability	3 (0.1751), 7 (0.2232), 10 (0.139)
	organizational challenges	9 (0.1117), 5 (0.1167), 9 (0.2094), 9 (0.1915), 9 (0.2775)
	business risks and costs	2 (0.3217), 9 (0.1996), 9 (0.14)
	sustainability in business models	3 (0.1149), 11 (0.2053), 1 (0.174)

For the below table, we didn't separate between items belonging to B/L and O/I. But items of O/I are remarked by bold font.

Table 43 Dominant topics for both kinds of issues

Aspect	Issue	Dominant topics (Contribution)
Technological	data resources	12 (0.1453), 10 (0.1038), 12 (0.0865), 7 (0.2025), 5 (0.0835), 12 (0.1174)
	data complexity	7 (0.1649), 12 (0.1494), 12 (0.0984), 7 (0.2407), 7 (0.1423), 10 (0.2132)
	limited infrastructures	12 (0.1921), 8 (0.1965)
	limited systems	12 (0.1162), 12 (0.1499)
	affecting technical solutions	4 (0.0777), 0 (0.0892), 11 (0.1556), 21 (0.1011), 11 (0.0871), 11 (0.1123), 9 (0.1082), 8 (0.0903), 8 (0.1349), 11 (0.1026), 7 (0.1547), 7 (0.1046), 9 (0.183)
Policy	public and privacy policy	13 (0.1017), 11 (0.1209), 11 (0.1202), 17 (0.097), 9 (0.0987), 13 (0.0837), 13 (0.2214), 13 (0.1866), 14 (0.1151)
Legal	data privacy and protection	19 (0.1281), 19 (0.2131), 1 (0.1346), 19 (0.1914), 13 (0.0684), 1 (0.1206), 16 (0.0788), 19 (0.1219), 10 (0.1003), 2 (0.1625), 19 (0.1074), 2 (0.1813), 18 (0.1446), 19 (0.0997), 12 (0.0841), 20 (0.1002), 4 (0.1112), 0 (0.0826), 15 (0.1529), 9 (0.1469), 11 (0.161), 13 (0.1533), 15 (0.1873), 11 (0.2104)
	(cyber-) security and breach-related obligations	15 (0.106), 0 (0.0815), 10 (0.0918), 2 (0.1576), 2 (0.1884), 2 (0.1444), 15 (0.2077), 2 (0.1514), 2 (0.1699), 13 (0.2118), 16 (0.1097), 1 (0.0704), 15 (0.2056), 15 (0.0726), 15 (0.2105)
	anonymization and pseudonymisation	2 (0.1325), 2 (0.1461), 4 (0.1081), 2 (0.1061), 2 (0.1029), 0 (0.0845), 19 (0.0909), 16 (0.1097), 1 (0.0704), 15 (0.2056), 15 (0.0726), 15 (0.2105)
	the supply of digital content and services	20 (0.154), 20 (0.1753), 20 (0.1772), 20 (0.2385), 19 (0.1014), 20 (0.1379), 20 (0.1858)
	the free flow of data	19 (0.1746), 17 (0.1275), 13 (0.1902), 13 (0.1582), 13 (0.1568), 21 (0.1268), 17 (0.0951), 21 (0.1552)
	intellectual property rights	16 (0.1287), 16 (0.1214), 20 (0.1687), 16 (0.1917), 13 (0.0996), 16 (0.1693), 16 (0.1523), 16 (0.1796), 0 (0.0877), 20 (0.1083), 14 (0.0995), 17 (0.0808), 17 (0.299), 17 (0.3231), 17 (0.1957)
	open data	13 (0.1922), 16 (0.1788), 16 (0.2082), 13 (0.0764), 13 (0.1586), 5 (0.1405), 5 (0.1426), 6 (0.0927)
	data sharing obligations	14 (0.1539), 14 (0.1542), 14 (0.0689), 14 (0.1232), 14 (0.132)
	data ownership-LEG	14 (0.1013), 17 (0.1504)

	data sharing agreements	21 (0.1002), 16 (0.1347), 14 (0.1204), 14 (0.0857), 14 (0.1558), 14 (0.1501), 14 (0.0911)
	liability	3 (0.0887), 20 (0.1401), 17 (0.1151)
	competition	18 (0.2165), 18 (0.0963), 14 (0.1039), 13 (0.0787), 17 (0.0908), 6 (0.2382)
Ethical and social	trust	1 (0.0801), 18 (0.075), 11 (0.1115), 12 (0.0898), 17 (0.0913), 8 (0.0842), 7 (0.0869)
	surveillance	11 (0.1013), 0 (0.1397), 0 (0.1094), 1 (0.0661), 10 (0.0874), 11 (0.1803), 18 (0.0903), 8 (0.0882), 21 (0.1003)
	privacy	11 (0.07), 5 (0.0847), 11 (0.0828), 15 (0.0743), 17 (0.0836), 0 (0.0981), 11 (0.1339)
	free will	12 (0.1332), 0 (0.0767), 9 (0.0831), 10 (0.0828), 21 (0.0967), 8 (0.1265)
	data ownership-ES	17 (0.0906), 19 (0.1414), 11 (0.0912), 1 (0.0604), 1 (0.0902), 11 (0.1634), 4 (0.1256)
	social discrimination	15 (0.0737), 8 (0.0909), 10 (0.097), 15 (0.0871), 10 (0.0909), 8 (0.0811)
	environment	10 (0.0907), 10 (0.1273), 6 (0.0973)
	others	4 (0.1347), 4 (0.1234), 4 (0.1593), 19 (0.1094), 4 (0.0935), 4 (0.1292)
Economic and political	inaccurate results	13 (0.0858), 9 (0.1334)
	restricted access to data	4 (0.1336), 11 (0.0931), 8 (0.1214)
	organizations lack capacity	5 (0.2776), 18 (0.0863), 9 (0.1263), 4 (0.1579), 5 (0.1316)
	big data industry underdeveloped	18 (0.1002), 0 (0.0881), 9 (0.1085)
	uncertain political risk	5 (0.1164)
	value creation	6 (0.101), 6 (0.2573), 12 (0.1847), 6 (0.0921), 8 (0.19), 10 (0.0961), 12 (0.1093), 6 (0.1708), 6 (0.1638), 15 (0.0697), 9 (0.1528), 8 (0.1782), 6 (0.1556), 5 (0.1039)
	data technology and systems	9 (0.1264)
	data availability	3 (0.1803), 12 (0.1619), 3 (0.2349)
	organizational challenges	3 (0.1272), 17 (0.0896), 3 (0.2271), 6 (0.1247), 5 (0.2049)
	business risks and costs	8 (0.1995), 5 (0.1002), 18 (0.2013)
	sustainability in business models	3 (0.0879), 9 (0.0729), 18 (0.1933)

The below table lists ID numbers and corresponding keywords of topics used in Table 41, Table 42 and Table 43.

Table 44 Representative keywords for extracted topics

	ID	Keywords
BL	0	transport, data_sharing, big_data_analytics, transport_sector, actor, dataset, sector, intellectual_property, public, intellectual
	1	difficult, application, may_prove, security, due, particularly_difficult, requirement, risk, identify, level
	2	big_data, big, analytic, may_lead, member_states, base, amount, notably, present, demonstrate
	3	datum, data, business, model, organization, algorithm, business_models, result, purpose, worker
	4	context, issue, big_data_context, practice, limit, case, account, hinder, rely, member
	5	protection, technology, require, copyright, database, analysis, legislation, copyright_protection, make, flow
	6	datum, information, anonymisation, company, protection_legislation, sensitive, measure, open, sensitive_information, integrate
	7	directive, lead, legal, legal_uncertainty, term, psi, exist, psi_directive, uncertainty, party

	8	personal_data, personal, process, relate, non_personal, apply, processing_activities, anonymisation_techniques, activity, become_personal
	9	user, privacy, ownership, time, include, collection, challenge, real_time, traffic, end_users
	10	datum, data_subjects, lack, source, subject, system, agreement, processing, size, affect
	11	service, digital, content, access, digital_content, obligation, current, provide, collect, remain
OI	0	improve, include, service, drive, data_value, service_providers, chain, cross_border, policy, development
	1	datum, personal_data, personal, level, require, processing, legal, eliminate, framework, relate
	2	big_data, service, big, infrastructure, cost, reduce, model, decision, cloud, offer
	3	business, management, network, time, data_collection, good, real_time, collection, vehicle, traffic
	4	privacy, issue, process, data_protection, protection, user, obligation, organisation, involve, personal_information
	5	datum, analytic, big_data_analytics, quality, big, lead, data_platforms, ensure, contractual, understand
	6	datum, data_sharing, share, research, industry, data, benefit, sharing, relevant, ownership
	7	provide, lemo_cs, information, application, solution, platform, customer, source, opportunity, lemo
	8	transport, transport_sector, create, sector, develop, requirement, transport_systems, case, set, address
	9	support, company, make, system, bda, change, risk, could_also, encourage, program
	10	public, open_data, open, access, adopt, start_ups, public_transport, standard, private_actors, private
	11	technology, increase, data_subjects, big, activity, subject, assessment, trust, opportunity, exist
BO	0	big_data, big, big_data_context, big_data_applications, specific, private_actors, possibility, pose, database, principle
	1	datum, issue, access, ownership, collect, due, relation, complex, become_personal, transaction
	2	application, difficult, security, anonymisation_techniques, may_prove, anonymisation, technique, particularly_difficult, prove, order
	3	develop, industry, research, practice, adopt, work, good, standard, encourage, contractual
	4	datum, data_subjects, subject, identify, lack, control, limit, worker, arrangement, affect
	5	company, lead, open, may_lead, information, organization, open_data, start, start_ups, sensitive
	6	transport, support, public, create, network, term, public_transport, bda, long, benefit
	7	datum, make, lemo_cs, information, platform, present, data_analytics, large, lemo, high
	8	service, provide, improve, increase, infrastructure, cost, customer, service_providers, provider, cloud
	9	technology, business, model, individual, development, business_model, business_models, transportation, environment, demand
	10	analytic, datum, big_data_analytics, big, case, study, case_studies, exist, condition, predictive
	11	privacy, user, relate, organisation, concern, give, people, trust, end_users, public_policies
	12	source, system, time, collection, quality, real_time, data_sources, traffic, real, management
	13	cross_border, policy, drive, situation, state, project, authority, psi_directive, border, policy_framework
	14	datum, data_sharing, share, sharing, data, technical, private, value_cycle, innovation, multiple
	15	include, actor, requirement, data, opportunity, value_chain, nis_directive, chain, compliance, stakeholder
	16	protection, context, directive, copyright, legislation, party, database, psi, protect, copyright_protection
	17	transport_sector, datum, sector, transport, intellectual, dataset, intellectual_property, property, part, behaviour
	18	require, risk, level, analysis, assessment, reduce, investment, solution, impact, account
	19	personal_data, personal, process, activity, processing_activities, non_personal, apply, law, extent, limitation
	20	legal, framework, content, digital, digital_content, current, vehicle, legal_uncertainty, obligation, legal_certainty
	21	datum, free, flow, free_flow, follow, uncertainty, agreement, decision_making, eliminate, travel

The below tables show the results for Figure 13, Figure 14 and Figure 15, respectively. Note that the column names were omitted since the columns are same with rows.

Table 45 Adjacency matrix of common items' number for BL issues

data resources	5	3	2	1	8	1	2	1	3	0	0	0	0	0	0	3	0	1	0	6	2	2	4	0	0	2	0	1	0	0	0
data complexity	3	3	1	1	4	3	2	2	2	0	3	3	1	1	1	3	0	1	0	3	1	1	3	0	0	2	0	1	0	0	0
limited infrastructures	2	1	1	0	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3	1	1	2	0	0	0	0	0	0	0	0
limited systems	1	1	0	1	0	1	2	1	1	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	2	0	1	0	0	0
affecting technical solutions	8	4	4	0	19	0	2	2	5	4	0	1	1	0	0	0	0	0	2	14	5	5	8	1	1	0	2	1	4	2	1
public and private policy	1	3	0	1	0	5	2	3	1	0	6	6	2	2	2	3	0	1	0	0	0	0	2	0	0	2	0	1	0	0	0
data privacy and protection	2	2	0	2	2	2	40	10	18	5	5	18	2	0	0	6	1	4	2	2	1	1	5	1	1	9	2	5	4	4	1
(cyber-) security and breach-related obligations	1	2	0	1	2	3	10	24	3	3	3	8	9	1	1	9	3	7	7	4	0	2	1	0	2	2	0	1	0	4	2
anonymization and pseudonymisation	3	2	1	1	5	1	18	3	13	3	3	1	1	0	1	3	0	1	2	5	1	2	5	2	1	5	0	1	1	2	1
the supply of digital content and services	0	0	0	0	4	0	5	3	3	18	1	1	2	0	0	2	1	1	1	0	4	0	1	0	0	1	0	0	0	0	0
the free flow of data	0	3	0	0	0	6	5	3	3	1	10	9	3	3	3	0	0	0	0	0	0	0	4	0	0	1	0	0	0	0	0
intellectual property rights	0	3	0	0	1	6	18	8	1	1	9	76	7	3	3	2	2	1	3	2	1	1	3	0	1	0	0	8	0	10	1
open data	0	1	0	0	1	2	2	9	1	2	3	7	7	1	1	4	3	2	4	2	1	1	1	0	1	0	0	0	0	2	1
data sharing obligations	0	1	0	0	0	2	0	1	0	0	3	3	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
data ownership-LEG	0	1	0	0	0	2	0	1	1	0	3	3	1	1	2	0	0	0	0	0	0	0	1	2	0	0	0	0	1	0	0
data sharing agreements	3	3	0	3	0	3	6	9	3	2	0	2	4	0	0	13	2	5	2	0	0	0	0	0	0	6	0	3	0	0	0
liability	0	0	0	0	0	0	1	3	0	1	0	2	3	0	0	2	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0
competition	1	1	0	1	0	1	4	7	1	1	0	1	2	0	0	5	1	3	1	0	0	0	0	0	0	2	0	1	0	0	0
trust	0	0	0	0	2	0	2	7	2	1	0	3	4	0	0	2	1	1	5	4	0	2	0	0	2	0	0	0	0	4	2
surveillance	6	3	3	0	14	0	2	4	5	0	0	2	2	0	0	0	0	0	4	13	3	5	6	0	2	0	0	0	0	4	2
privacy	2	1	1	0	5	0	1	0	1	4	0	1	1	0	0	0	1	0	0	3	3	1	2	0	0	0	0	0	0	0	0
free will	2	1	1	0	5	0	1	2	2	0	0	1	1	0	0	0	0	0	2	5	1	2	2	0	1	0	0	0	0	2	1
data ownership-ES	4	3	2	0	8	2	5	1	5	1	4	3	1	1	1	0	0	0	0	6	2	2	6	0	0	1	0	0	0	0	0

social discrimination	0	0	0	0	1	0	1	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	5	0	0	2	1	6	0	0
environment	0	0	0	0	1	0	1	2	1	0	0	1	1	0	0	0	0	0	2	2	0	1	0	0	1	0	0	0	0	2	1
others	2	2	0	2	0	2	9	2	5	1	1	0	0	0	0	6	0	2	0	0	0	0	1	0	0	5	0	2	0	0	0
inaccurate results	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	4	2	8	0	0
restricted access to data	1	1	0	1	1	1	5	1	1	0	0	8	0	0	0	3	0	1	0	0	0	0	0	1	0	2	2	3	4	1	0
organizations lack capacity	0	0	0	0	4	0	4	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	6	0	0	8	4	17	0	0
big data industry underdeveloped	0	0	0	0	2	0	4	4	2	0	0	10	2	0	0	0	0	0	4	4	0	2	0	0	2	0	0	1	0	5	2
uncertain political risk	0	0	0	0	1	0	1	2	1	0	0	1	1	0	0	0	0	0	2	2	0	1	0	0	1	0	0	0	0	2	1

Table 46 Adjacency matrix of common items' number for OI issues

data resources	3	1	0	1	3	0	1	0	0	0	0	0	0	1	2	0	0	1	0	0	2	1	3	1	0	2	0	1	1	0	0
data complexity	1	3	1	0	3	0	1	0	0	0	0	0	0	0	0	0	1	2	1	1	3	0	2	1	0	7	1	1	1	1	1
limited infrastructures	0	1	1	0	2	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	0	0	0	1	0
limited systems	1	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
affecting technical solutions	3	3	2	3	14	0	2	0	0	0	0	0	0	0	0	0	1	2	2	2	2	0	0	0	0	7	1	4	0	2	1
public and private policy	0	0	0	0	0	12	9	10	0	0	6	1	3	3	0	1	2	2	0	0	1	0	0	1	2	8	0	1	0	0	1
data privacy and protection	1	1	1	0	2	9	24	9	12	0	2	0	0	1	2	0	0	1	7	4	4	4	1	0	6	9	0	0	0	1	3
(cyber-) security and breach-related obligations	0	0	0	0	0	10	9	11	4	0	4	0	3	2	0	0	1	1	2	1	0	1	0	1	0	7	0	1	0	0	0
anonymization and pseudonymisation	0	0	0	0	0	0	12	4	16	0	0	0	0	0	0	0	0	0	0	8	4	0	4	0	0	0	4	0	0	0	0	0
the supply of digital content and services	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	1	0	0	0	0	0	1
the free flow of data	0	0	0	0	0	6	2	4	0	0	6	2	3	4	0	2	3	3	0	0	0	0	0	1	0	6	0	1	0	0	0
intellectual property rights	0	0	0	0	0	1	0	0	0	0	2	1	0	1	0	1	1	1	0	0	0	0	0	0	0	2	0	0	0	0	0
open data	0	0	0	0	0	3	0	3	0	0	3	0	9	6	0	0	3	3	0	0	0	0	0	3	0	0	0	3	0	0	0
data sharing obligations	1	0	0	0	0	3	1	2	0	0	4	1	6	6	2	1	3	3	0	0	0	1	1	2	0	3	0	2	0	0	0
data sharing agreements	2	0	0	0	0	0	2	0	0	0	0	0	0	2	4	0	0	0	0	0	0	2	2	0	0	2	0	0	0	0	0
liability	0	0	0	0	0	1	0	0	0	0	2	1	0	1	0	1	1	1	0	0	0	0	0	0	0	2	0	0	0	0	0
competition	0	1	0	0	1	2	0	1	0	0	3	1	3	3	0	1	3	2	0	0	0	0	0	1	0	7	1	2	0	0	1

trust	1	2	1	0	2	2	1	1	0	0	3	1	3	3	0	1	2	4	1	1	3	0	2	2	0	4	0	1	1	1	0
surveillance	0	1	1	0	2	0	7	2	8	0	0	0	0	0	0	0	0	1	6	3	1	2	0	0	0	4	0	0	4	3	0
privacy	0	1	1	0	2	0	4	1	4	2	0	0	0	0	0	0	0	1	3	6	1	3	0	0	2	2	0	0	0	1	2
free will	2	3	1	0	2	1	4	0	0	0	0	0	0	0	0	0	0	3	1	1	6	0	4	2	2	3	0	0	2	1	1
data ownership-ES	1	0	0	0	0	0	4	1	4	1	0	0	0	1	2	0	0	0	2	3	0	3	1	0	1	2	0	0	0	0	1
social discrimination	3	2	0	0	0	0	1	0	0	0	0	0	0	1	2	0	0	2	0	0	4	1	5	2	0	3	0	0	2	0	0
environment	1	1	0	0	0	1	0	1	0	0	1	0	3	2	0	0	1	2	0	0	2	0	2	2	0	1	0	1	1	0	0
others	0	0	0	0	0	2	6	0	0	1	0	0	0	0	0	0	0	0	0	2	2	1	0	0	5	0	0	0	0	0	3
value creation	2	7	1	0	7	8	9	7	4	0	6	2	0	3	2	2	7	4	4	2	3	2	3	1	0	38	5	5	5	3	5
data technology and systems	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	5	1	1	0	0	1
data availability	1	1	0	1	4	1	0	1	0	0	1	0	3	2	0	0	2	1	0	0	0	0	0	1	0	5	1	3	0	0	1
organizational challenges	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	0	2	0	2	1	0	5	0	0	17	8	0
business risks and costs	0	1	1	0	2	0	1	0	0	0	0	0	0	0	0	0	0	1	3	1	1	0	0	0	0	3	0	0	8	5	0
sustainability in business models	0	1	0	0	1	1	3	0	0	1	0	0	0	0	0	0	1	0	0	2	1	1	0	0	3	5	1	1	0	0	3

Table 47 Adjacency matrix of common items' number for BO issues

data resources	12	10	3	6	2	0	4	1	0	0	0	0	2	0	0	0	0	0	4	1	1	4	0	2	2	0	0	0	2	0	1	8	0	3	1	1	0
data complexity	10	14	2	4	6	0	3	1	0	0	0	0	0	0	0	0	0	0	5	1	0	3	0	2	2	0	0	0	0	0	0	5	0	2	0	0	0
limited infrastructures	3	2	2	2	2	0	1	0	0	0	0	0	0	0	0	0	0	0	2	1	0	2	0	2	0	0	0	1	0	0	0	4	0	1	0	1	0
limited systems	6	4	2	4	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	4	0	2	0	0	0
affecting technical solutions	2	6	2	0	31	10	12	1	2	0	2	1	0	0	0	1	0	0	8	13	13	6	9	4	0	5	2	7	3	3	0	6	2	0	0	2	2
public and private policy	0	0	0	0	10	23	13	4	1	0	14	9	12	5	2	5	1	6	3	4	7	1	5	0	0	0	5	2	1	1	0	1	1	0	1	0	1
data privacy and protection	4	3	1	2	12	13	64	27	26	12	12	11	8	0	0	1	1	4	6	10	9	4	15	6	2	11	3	3	3	3	0	6	1	1	0	1	2
(cyber-) security and breach-related obligations	1	1	0	0	1	4	27	55	31	0	3	8	5	0	0	1	0	1	1	4	6	2	2	12	2	0	1	0	0	1	0	6	0	0	0	0	0
anonymization and pseudonymisation	0	0	0	0	2	1	26	31	43	2	2	2	0	5	1	5	0	1	0	2	1	1	3	0	0	7	0	1	1	1	0	0	0	0	0	0	0

the supply of digital content and services	0	0	0	0	0	0	12	0	2	37	1	12	0	0	0	0	6	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
the free flow of data	0	0	0	0	2	14	12	3	2	1	18	11	9	0	2	2	2	5	2	2	2	2	3	0	0	1	3	0	0	0	0	0	0	0	2	0	0
intellectual property rights	0	0	0	0	1	9	11	8	2	12	11	59	15	5	5	11	6	6	4	2	5	1	4	0	0	0	1	0	0	1	0	0	0	0	4	0	0
open data	2	0	0	0	0	12	8	5	0	0	9	15	18	0	0	2	0	4	0	0	2	0	0	0	1	0	3	0	4	0	2	8	0	0	3	2	0
data sharing obligations	0	0	0	0	0	5	0	0	5	0	0	5	0	25	5	25	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
data ownership-LEG	0	0	0	0	0	2	0	0	1	0	2	5	0	5	2	5	1	2	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
data sharing agreements	0	0	0	0	1	5	1	1	5	0	2	11	2	25	5	27	0	5	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
liability	0	0	0	0	0	1	1	0	0	6	2	6	0	0	1	0	3	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2	3	0	1
competition	0	0	0	0	0	6	4	1	1	0	5	6	4	5	2	5	1	8	3	2	1	0	1	0	1	0	1	0	2	2	0	6	0	0	2	2	2
trust	4	5	2	2	8	3	6	1	0	0	2	4	0	0	1	0	1	3	7	5	4	2	5	2	0	0	0	2	1	1	0	4	0	1	1	2	1
surveillance	1	1	1	0	13	4	10	4	2	0	2	2	0	0	0	1	0	2	5	13	8	5	6	4	2	0	0	3	1	3	0	3	0	0	0	2	1
privacy	1	0	0	0	13	7	9	6	1	0	2	5	2	0	1	0	1	1	4	8	13	1	7	2	0	0	0	3	2	1	1	2	0	0	2	1	0
free will	4	3	2	2	6	1	4	2	1	0	2	1	0	0	0	1	0	0	2	5	1	6	0	4	2	0	1	1	1	2	0	6	1	1	0	1	1
data ownership-ES	0	0	0	0	9	5	15	2	3	1	3	4	0	0	1	0	1	1	5	6	7	0	11	0	0	6	0	3	1	0	0	0	0	0	1	0	0
social discrimination	2	2	2	0	4	0	6	12	0	0	0	0	0	0	0	0	0	0	2	4	2	4	0	12	4	0	0	2	0	0	0	8	0	0	0	2	0
environment	2	2	0	0	0	0	2	2	0	0	0	0	1	0	0	0	0	1	0	2	0	2	0	4	5	0	0	0	0	0	0	8	0	0	1	0	0
others	0	0	0	0	5	0	11	0	7	1	1	0	0	0	0	0	0	0	0	0	0	0	6	0	0	26	0	5	5	0	0	0	0	0	0	0	0
inaccurate results	0	0	0	0	2	5	3	1	0	0	3	1	3	0	0	0	0	1	0	0	0	1	0	0	0	0	2	0	1	1	0	1	1	0	0	0	1
restricted access to data	0	0	1	0	7	2	3	0	1	0	0	0	0	0	0	0	0	0	2	3	3	1	3	2	0	5	0	3	1	0	0	2	0	0	0	1	0
organizations lack capacity	2	0	0	0	3	1	3	0	1	0	0	0	4	0	0	0	0	2	1	1	2	1	1	0	0	5	1	1	7	2	2	3	1	0	2	3	2
big data industry underdeveloped	0	0	0	0	3	1	3	1	1	0	0	1	0	0	0	0	0	2	1	3	1	2	0	0	0	0	1	0	2	3	0	1	1	0	0	1	2
uncertain political risk	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	1	1	0	0	1	1	0
value creation	8	5	4	4	6	1	6	6	0	0	0	0	8	0	0	0	0	6	4	3	2	6	0	8	8	0	1	2	3	1	1	48	1	2	7	3	1
data technology and systems	0	0	0	0	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	1	0	0	0	1
data availability	3	2	1	2	0	0	1	0	0	0	0	0	0	0	0	0	2	0	1	0	0	1	0	0	0	0	0	0	0	0	0	2	0	5	4	0	2

organizational challenges	1	0	0	0	0	1	0	0	0	0	2	4	3	0	1	0	3	2	1	0	2	0	1	0	1	0	0	0	2	0	1	7	0	4	7	1	2
business risks and costs	1	0	1	0	2	0	1	0	0	0	0	0	2	0	0	0	0	2	2	2	1	1	0	2	0	0	0	1	3	1	1	3	0	0	1	3	1
sustainability in business models	0	0	0	0	2	1	2	0	0	0	0	0	0	0	0	0	1	2	1	1	0	1	0	0	0	0	1	0	2	2	0	1	1	2	2	1	3

Table 48 All processes for calculating severity and intensity

(1) with others (2) with others in the same aspects (3) with others in the other aspects	# of (1)	# of (2)	# of (3)	# of items	Normalised # of (1)	Normalized # of (2)	Normalised # of (3)	C for (1)	C for (2)	C for (3)	D for (1)	D for (2)	D for (3)	Rescaled value for (1)	Rescaled value for (2)	Rescaled value for (3)
BL-data resources	47	25	22	3	15.67	8.33	7.33	0.27	0.27	0.29	-0.76	0.96	0.71	3.44	3.21	2.81
OI-data resources	34	18	16	3	11.33	6.00	5.33	0.27	0.27	0.29	-0.76	0.96	0.71	2.28	2.58	2.24
BL-data complexity	44	19	25	3	14.67	6.33	8.33	0.27	0.27	0.29	-0.76	0.96	0.71	3.17	2.67	3.10
OI-data complexity	44	28	16	3	14.67	9.33	5.33	0.27	0.27	0.29	-0.76	0.96	0.71	3.17	3.47	2.24
BL-limited infrastructures	16	9	7	1	16.00	9.00	7.00	0.27	0.27	0.29	-0.76	0.96	0.71	3.53	3.38	2.71
OI-limited infrastructures	12	10	2	1	12.00	10.00	2.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.45	3.65	1.29
BL-limited systems	16	9	7	1	16.00	9.00	7.00	0.27	0.27	0.29	-0.76	0.96	0.71	3.53	3.38	2.71
OI-limited systems	16	9	7	1	16.00	9.00	7.00	0.27	0.27	0.29	-0.76	0.96	0.71	3.53	3.38	2.71
BL-affecting technical solutions	81	15	66	7	11.57	2.14	9.43	0.27	0.27	0.29	-0.76	0.96	0.71	2.34	1.54	3.41
OI-affecting technical solutions	64	28	36	6	10.67	4.67	6.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.10	2.22	2.43
BL-public and private policy	55	10	45	3	18.33	3.33	15.00	0.27	0.27	0.29	-0.76	0.96	0.71	4.15	1.86	5.00
OI-public and private policy	80	10	70	6	13.33	1.67	11.67	0.27	0.27	0.29	-0.76	0.96	0.71	2.81	1.41	4.05
BL-data privacy and protection	163	121	42	14	11.64	8.64	3.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.36	3.29	1.57
OI-data privacy and protection	111	56	55	10	11.10	5.60	5.50	0.27	0.27	0.29	-0.76	0.96	0.71	2.21	2.47	2.29
BL-(cyber-) security and breach-related obligations	105	73	32	10	10.50	7.30	3.20	0.27	0.27	0.29	-0.76	0.96	0.71	2.05	2.93	1.63
OI-(cyber-) security and breach-related obligations	51	43	8	5	10.20	8.60	1.60	0.27	0.27	0.29	-0.76	0.96	0.71	1.97	3.28	1.17

BL-anonymization and pseudonymisation	75	32	43	7	10.71	4.57	6.14	0.27	0.27	0.29	-0.76	0.96	0.71	2.11	2.19	2.47
OI-anonymization and pseudonymisation	58	44	14	4	14.50	11.00	3.50	0.27	0.27	0.29	-0.76	0.96	0.71	3.12	3.92	1.71
BL-the supply of digital content and services	58	44	14	6	9.67	7.33	2.33	0.27	0.27	0.29	-0.76	0.96	0.71	1.83	2.94	1.38
OI-the supply of digital content and services	13	11	2	1	13.00	11.00	2.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.72	3.92	1.29
BL-the free flow of data	38	28	10	4	9.50	7.00	2.50	0.27	0.27	0.29	-0.76	0.96	0.71	1.78	2.85	1.43
OI-the free flow of data	36	26	10	4	9.00	6.50	2.50	0.27	0.27	0.29	-0.76	0.96	0.71	1.65	2.71	1.43
BL-intellectual property rights	106	68	38	14	7.57	4.86	2.71	0.27	0.27	0.29	-0.76	0.96	0.71	1.27	2.27	1.49
OI-intellectual property rights	11	8	3	1	11.00	8.00	3.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.19	3.12	1.57
BL-open data	45	35	10	5	9.00	7.00	2.00	0.27	0.27	0.29	-0.76	0.96	0.71	1.65	2.85	1.29
OI-open data	32	11	21	2	16.00	5.50	10.50	0.27	0.27	0.29	-0.76	0.96	0.71	3.53	2.44	3.71
BL-data sharing obligations	17	11	6	1	17.00	11.00	6.00	0.27	0.27	0.29	-0.76	0.96	0.71	3.79	3.92	2.43
OI-data sharing obligations	61	28	33	4	15.25	7.00	8.25	0.27	0.27	0.29	-0.76	0.96	0.71	3.33	2.85	3.07
BL-data ownership-LEG	26	19	7	2	13.00	9.50	3.50	0.27	0.27	0.29	-0.76	0.96	0.71	2.72	3.52	1.71
BL-data sharing agreements	66	36	30	5	13.20	7.20	6.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.78	2.90	2.43
OI-data sharing agreements	32	20	12	2	16.00	10.00	6.00	0.27	0.27	0.29	-0.76	0.96	0.71	3.53	3.65	2.43
BL-liability	20	18	2	2	10.00	9.00	1.00	0.27	0.27	0.29	-0.76	0.96	0.71	1.92	3.38	1.00
OI-liability	12	11	1	1	12.00	11.00	1.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.45	3.92	1.00
BL-competition	26	6	20	3	8.67	2.00	6.67	0.27	0.27	0.29	-0.76	0.96	0.71	1.56	1.50	2.62
OI-competition	25	13	12	3	8.33	4.33	4.00	0.27	0.27	0.29	-0.76	0.96	0.71	1.47	2.13	1.86
BL-trust	24	18	6	3	8.00	6.00	2.00	0.27	0.27	0.29	-0.76	0.96	0.71	1.38	2.58	1.29
OI-trust	41	15	26	4	10.25	3.75	6.50	0.27	0.27	0.29	-0.76	0.96	0.71	1.99	1.97	2.57
BL-surveillance	44	26	18	5	8.80	5.20	3.60	0.27	0.27	0.29	-0.76	0.96	0.71	1.60	2.36	1.74
OI-surveillance	43	14	29	4	10.75	3.50	7.25	0.27	0.27	0.29	-0.76	0.96	0.71	2.12	1.90	2.79
BL-privacy	32	12	20	3	10.67	4.00	6.67	0.27	0.27	0.29	-0.76	0.96	0.71	2.10	2.04	2.62
OI-privacy	35	16	19	4	8.75	4.00	4.75	0.27	0.27	0.29	-0.76	0.96	0.71	1.58	2.04	2.07
BL-free will	28	10	18	2	14.00	5.00	9.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.99	2.31	3.29

OI-free will	44	14	30	4	11.00	3.50	7.50	0.27	0.27	0.29	-0.76	0.96	0.71	2.19	1.90	2.86
BL-data ownership-ES	45	11	34	4	11.25	2.75	8.50	0.27	0.27	0.29	-0.76	0.96	0.71	2.25	1.70	3.14
OI-data ownership-ES	37	13	24	3	12.33	4.33	8.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.54	2.13	3.00
BL-social discrimination	28	9	19	3	9.33	3.00	6.33	0.27	0.27	0.29	-0.76	0.96	0.71	1.74	1.77	2.52
OI-social discrimination	28	9	19	3	9.33	3.00	6.33	0.27	0.27	0.29	-0.76	0.96	0.71	1.74	1.77	2.52
BL-environment	14	8	6	1	14.00	8.00	6.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.99	3.12	2.43
OI-environment	26	5	21	2	13.00	2.50	10.50	0.27	0.27	0.29	-0.76	0.96	0.71	2.72	1.63	3.71
BL-others	33	18	15	3	11.00	6.00	5.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.19	2.58	2.14
OI-others	33	18	15	3	11.00	6.00	5.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.19	2.58	2.14
BL-inaccurate results	43	30	13	2	21.50	15.00	6.50	0.27	0.27	0.29	-0.76	0.96	0.71	5.00	5.00	2.57
BL-restricted access to data	44	22	22	3	14.67	7.33	7.33	0.27	0.27	0.29	-0.76	0.96	0.71	3.17	2.94	2.81
BL-organizations lack capacity	78	44	34	5	15.60	8.80	6.80	0.27	0.27	0.29	-0.76	0.96	0.71	3.42	3.33	2.66
BL-big data industry underdeveloped	46	22	24	3	15.33	7.33	8.00	0.27	0.27	0.29	-0.76	0.96	0.71	3.35	2.94	3.00
BL-uncertain political risk	17	10	7	1	17.00	10.00	7.00	0.27	0.27	0.29	-0.76	0.96	0.71	3.79	3.65	2.71
OI-value creation	92	2	90	14	6.57	0.14	6.43	0.27	0.27	0.29	-0.76	0.96	0.71	1.00	1.00	2.55
OI-data technology and systems	20	14	6	1	20.00	14.00	6.00	0.27	0.27	0.29	-0.76	0.96	0.71	4.60	4.73	2.43
OI-data availability	59	28	31	3	19.67	9.33	10.33	0.27	0.27	0.29	-0.76	0.96	0.71	4.51	3.47	3.67
OI-organizational challenges	65	30	35	5	13.00	6.00	7.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.72	2.58	2.71
OI-business risks and costs	40	25	15	3	13.33	8.33	5.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.81	3.21	2.14
OI-sustainability in business models	33	15	18	3	11.00	5.00	6.00	0.27	0.27	0.29	-0.76	0.96	0.71	2.19	2.31	2.43