MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

D5.2 Report on Quantification and Measurement of the Data

19/12/2022

Author(s): Selini HADJIDIMITRIOU (ICOOR), Piergiuseppe DI GREGORIO (ICOOR), Giulia RENZI (ICOOR)

MobiDataLab is funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879).

Summary sheet

Deliverable Number	D5.2
Deliverable Name	D5.2 - Report on Quantification and Measurement of the Data Exchange Culture
Full Project Title	MobiDataLab, Labs for prototyping future Mobility Data sharing cloud solutions
Responsible Author(s)	Selini HADJIDIMITRIOU (ICOOR)
Contributing Partner(s)	AETHON, POLIS, HERE, AKKA
Peer Review	Diana GUARDADO (FS6), Danijel PAVLICA (F6S),Renée OBREGON (AKKA), Didier DE RYCK (KISIO)
Contractual Delivery Date	31-10-2022
Actual Delivery Date	28-10-2022
Status	Final
Dissemination level	Public
Version	V1.0
No. of Pages	54
WP/Task related to the deliverable	WP5/T5.2
WP/Task responsible	ICOOR
Document ID	MobiDataLab-D5.2-ReportQuantificationMeasurementDataExchangeCulture-v1.0
Abstract	This deliverable describes the preliminary analysis of the data-sharing culture.

Legal Disclaimer

Project partners

Organisation	Country	Abbreviation
CONSORZIO UNIVERSITARIO PER L'OTTIMIZZAZIONE E LA RICERCA OPERATIVA	Italy	ICOOR
AETHON SYMVOULI MICHANIKI MONOPROSOPI IKE	Greece	AETHON
POLIS - PROMOTION OF OPERATIONAL LINKS WITH INTEGRATED SERVICES, ASSOCIATION INTERNATIONALE	Belgium	POLIS
HERE GLOBAL B.V.	Germany	HERE
AKKA RESEARCH	France	AKKA

Document history

Version	Date	Organisation	Main area of changes	Comments
0.1	01/06/2022	ICOOR	All	Creating the first draft
0.2	15/07/2022	ICOOR	Chapter 2	Analysis of the literature
0.3	15/09/2022	ICOOR	Chapter 3, 4 and 5	Topic modelling, analysis of survey results
0.4	22/10/2022	ICOOR	All	Final draft of the deliverable
0.5	25/10/2022	ICOOR	First page	Editing of the first page
0.6	26/10/2022	AKKA	All	1st QC
0.7	27/10/2022	F6S-KISIO-AKKA	All	Peer-Review
0.8	28/10/2022	ICOOR	All	Consolidation
0.9	28/10/2022	AKKA	All	2nd QC
1.0	28/10/2022	AKKA	All	Submission of Final Version

Executive Summary

This document presents the preliminary results of the data-sharing culture's quantification and measurement. More specifically, Chapter 1 analyses the existing literature on argumentation mining, topic modelling, business models' definition and the theory of acceptance of technologies. Chapter 3 describes the interviews and the six topics identified thanks to argumentation mining and topic modelling. The topics are: Prerequisites for data sharing (Topic I and II), Social impact of data sharing (Topic III), Challenges (Topic IV), Ecosystems of data sharing (Topic V) and Vision (Topic VI). The Chapter discusses each topic based on the existing literature and reports relevant citations of the interviewees. Chapter 4 analyses the results of the 39 interviews to assess the acceptance of the Transport Cloud. According to the received answers, the respondents are willing to use the Transport Cloud if it is easy, has a fast response time and ensures data security. Chapter 5 presents the methodology to evaluate the business models' acceptance. The analysis considers a sample of 36 data-sharing companies identified in D3.2 Data sharing market technological developments monitoring and describes their characteristics based on the Canvas categories. The agglomerative cluster analysis identifies six clusters of companies to propose a new classification of business models of data-sharing companies. Finally, Chapter 6 presents the future work in the context of T5.2 Quantification and measurement of the data exchange culture to perform additional interviews, collecting answers to assess the Transport Cloud acceptance and refining the methodology on the business models acceptance.

Funded by the European Union

Table of contents

List of figures

List of tables

Abbreviations and acronyms

Abbreviation	Meaning
BM	Business Models
PaaS	Platform as a Service
IaaS	Infrastructure as a Service
TAM	Technology Acceptance Model
UTAUT	Unified Theory of Acceptance and Use of Technology
LDA	Latent Dirichlet Allocation
AM	Argumentation Mining
NLP	Natural Language Processing
OEM	Original Equipment Manufacturer

Funded by the European Union

1. Introduction

This deliverable presents a preliminary implementation of the evaluation methodology presented in D3.5 Societal and Environmental Impacts of Data Sharing assessment framework, D5.2 Quantification and measurement of the data exchange culture describes the data collected and how these data will be elaborated and presented in D5.3 Analysis and conclusions on the data exchange culture. Based on this preliminary data collection, it was possible to refine the evaluation methodology proposed in D3.5 Societal and Environmental Impacts of Data Sharing assessment framework". The first three interviews on the data sharing culture have allowed for revising the protocol for the interview. D5.2 Quantification and measurement of the data exchange culture presents the preliminary results of topic modelling and argumentation mining. Topic modelling is a Natural Language Processing (NLP) technique that allows the identification of topics in a corpus. Similarly, argumentation mining extract claims and evidence from a corpus. Furthermore, the document shows how to identify the main features of data-sharing companies based on their business models. Finally, D5.2 Quantification and measurement of the data exchange culture describes the preliminary results of the Transport Cloud acceptance survey. The last part of the deliverable describes the future activities to finalize the analysis, refine the methodologies and improve data collection processes.

1.1. Purpose of the document and target group

This deliverable presents the preliminary results of the qualitative and quantitative analyses proposed in D3.5 Societal and Environmental Impacts of Data Sharing assessment framework. The first analysis focuses on the data exchange culture to understand the experience of the different organizations on data sharing. The interviews with several stakeholders gather the point of view on data-sharing practices within and among public and private organizations. Furthermore, the objective of the interviews is to understand what it means data quality and assess the social and environmental impact of the data-sharing culture. Several stakeholders can be interested in the results of this analysis, such as the MobiDataLab partners and the interviewed actors. Data sharing and the impact this might have at organizational and societal levels have high interest for all the different interviewed stakeholders.

The second objective of the deliverable is to present the preliminary results of the evaluation of the business model's acceptance. The proposed methodology aims to highlight the main characteristics of the data-sharing companies concerning the business model that they have selected. Since the objective is to highlight the main features of the companies according to their business model, the preliminary analysis mainly focuses on private companies. For this reason, besides the MobiDataLab project participants, the target audience could be private companies interested in providing data-sharing services and could consider different business models.

Finally, the last preliminary analysis of the deliverable concerns the Transport Cloud acceptance based on the Unified Theory of Acceptance and Use of Technology (UTAUT). Between June and October 2022, 39 respondents participated in the online survey proposed in D3.5. It turned out that the respondents who were not part of the MobidataLab project found it hard to understand what the

Funded by the European Union

Transport Cloud was. The description of the Transport Cloud will be improved during the next few months by the MobiDataLab partners who are in charge of developing the prototype to gather additional answers during the events organized by MobiDataLab. The target group of the results of the Transport Cloud acceptance is the project participants, especially those in charge of developing the MobiDataLab Transport Cloud prototype.

1.2. Contribution of partners

ICOOR is the leader of the T5.2 Quantification and measurement of the data exchange culture and the responsible partner of D5.2 Report on Quantification and measurement of the data exchange culture. ICOOR has carried out the data collection, the preliminary data analysis, and the drafting of the deliverable. AETHON has strongly supported data collection for the online survey on Transport Cloud acceptance.

Three partners have contributed to the D5.2 Report on Quantification and measurement of the data exchange culture. POLIS has contacted several actors to participate in the interview on the data-sharing culture. Currently, ICOOR interviewed eleven actors, and the plan is to have forty interviews by the end of the MobiDataLab project. ICOOR has already interviewed HERE, POLIS and AKKA. Furthermore, HERE has contributed by commenting on the online survey on the Transport Cloud acceptance, especially on what concerns the description of the cloud solution that helps the respondents who are not participating in the MobidataLab project to participate in the online survey. AKKA has described the Transport Cloud to explain it to the respondents who are not part of MobiDataLab.

1.3. Relation to other activities

T5.2 aims to quantify and measure the data-sharing culture. D5.2 Report on Quantification and measurement of the data exchange culture concerns the preliminary analysis based on the methodology described in D3.5 Societal and Environmental Impacts of Data Sharing assessment framework. D5.2 Report on Quantification and measurement of the data exchange culture is related to D3.1 Actors' needs cooperation framework because the interviews are organized by considering the categories of actors identified in that document. Furthermore, the acceptance of the business models is evaluated based on the elaboration of data reported in D3.2 Data sharing market technological developments monitoring and D3.4 Data sharing business and revenue models in data sharing. Finally, D5.2 Report on Quantification and measurement of the data exchange culture is related to D5.1 Virtual lab because the deliverable includes a set of KPIs that could be important for the Transport Cloud acceptance evaluation.

2. Literature review

2.1. Topic modelling

Funded by the European Union

Topic modelling is a statistical method that allows theme detection based on the analysis of words used in the text to understand how topics are connected and change over time (Blei, 2012). Topic modelling for the analysis of contents has been utilized to identify the main themes in surveys (Baumer et al., 2017), and to analyse high dimension text data (Liu and Xu, 2017; Seshadri, Mercy Shalinie and Kollengode, 2015; Song et al., 2016; Zhao, Jin and Yue, 2015). Concerning the difference between the manual analysis of experts and topic modelling, Baumer et al. (2017) pointed out that the extracted themes were similar. Suominen and Toivanen (2016) compared the automated classification of text to manual analysis to classify scientific works in Finland. They found that each method has advantages and disadvantages depending on the objectives.

Latent Dirichlet Allocation (LDA) is a technique commonly deployed to perform topic modelling. According to LDA, hidden variables generate topics. Therefore, the approach consists of computing a joint distribution of hidden variables based on the observed ones (Blei, 2012). Each document contributes to the topic to a certain degree, and each topic generates words of the vocabulary. The input of the LDA is a matrix in which the rows are the documents, and the columns are the number of terms included in each document (document-term matrix). The output of the LDA are two matrices: document-topic distribution and topic-word distribution. The document-topic distribution reports the probability distribution of the topics present in each article. The topic-words distribution is a matrix that reports the probability of the distribution of words generated by each topic.

Table 1: Document-term matrix

	Word 1	Word 2	...	Word N
Document 1	0	0	...	1
Document 2	2	2	...	0
Document 3	3	5	...	1
…	…	…	...	...
Document M	0	0	...	0

Table 2: Document-topic distribution

	Topic 1	Topic 2	...	Topic K
Document 1	P(t1|a1)	P(t2|a1)	...	P(tK|a1)
Document 2	P(t1|a2)	P(t2|a2)	...	P(tK|a2)
Document 3	P(t1|a3)	P(t2|a3)	...	P(tK|a3)
…	…	…	...	...

Funded by the European Union

Document M	P(t1|aM)	P(t2|aM)	...	P(tK|aM)

Table 3: Topic-word distribution

	Word 1	Word 2	...	Word N
Topic 1	P(w1|t1)	P(w2|t1)	...	P(wN|t1)
Topic 2	P(w1|t2)	P(w2|t2)	...	P(wN|t2)
Topic 3	P(w1|t3)	P(w2|t3)	...	P(wN|t3)
…	…	…	...	...
Topic M	P(w1|tM)	P(w2|tM)	...	P(wN|tM)

Other challenges of applying topic analysis for context analysis deal with the correct number of topics and, most importantly, how to validate the results. The first step is to stem words so that similar words have the same root. Stemming allows for assessing the words' frequency and improves the quality of the results (Hopkins and King, 2010). The most popular stemming algorithm is the Porter stemmer. According to Hagen et al. (2015), full stemming is preferable to no stemming or minimal stemming that only standardises verb declination. Concerning model validation, perplexity can help to estimate the optimal number of topics. However, perplexity mainly measures the model adaptation. In this regard, Chang et al. (2009) found a negative correlation between the model and manual analysis (experts' human judgment) perplexity. Furthermore, experts' human judgment is often unpractical, especially for large volumes of texts. For this reason, Hange (2009) proposed a two-step process to obtain more accurate results using perplexity. In the first step, the model's performance is evaluated using perplexity and several topics in a cross-validation framework. The second step assesses the quality of the identified themes.

In the literature, there are many topic modelling methods and tools. Barde and Bainwad (2017) described four approaches to implement topic modelling and their corresponding limitations: Vector Space Modeling (VSM), Latent Semantic Indexing (LSI), Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA). Moreover, the authors reviewed some tools to implement topic modelling: Gensim and Stanford Topic Modeling Toolbox (TMT).

Word2Vec is a technique introduced by Mikolov et al. (2013) that allows embedding words and learning their similarities using neural networks. Vectors represent words, and the similarity of words depends on their proximity. The main idea is that the meaning of words is inferred based on their most similar words. For instance, “beautiful”, “cute” and “pretty” are related and can appear in a similar context. Word2vec can find similar or unrelated topics and measure the similarity between couples of words. There are two approaches to implementing Word2Vec: Skip Gram and Common Bag of Words (CBOW).

Funded by the European Union

2.2. Argumentation Mining

Argumentation mining is a multi-disciplinary domain that analyses debate and reasoning processes (Lippi &Torroni, 2016). These studies have gained increasing importance in recent years due to new machine learning and computational linguistic approaches. Argumentation mining (AM) represents the set of techniques for automatically extrapolating topics from generic texts (Lippi & Torroni, 2016). The methodology is gaining importance thanks to the wide range of potential applications, such as web tools, user-generated content related to business analyses, and legal documents. The core of the AM approach is the argument, defined as a set of statements formed by three parts (Lippi & Torroni, 2016): premises, inference from the premises, and the conclusion. The challenge of Argumentation Mining is breaking the arguments into sections, detecting the argument components, and identifying the claims. Since there is no standardization in the corpora, the literature presents several models aimed at extrapolating the topics and the evidence from the text. Some models, such as IBM, start from a given topic and proceed to find the evidence, hence knowing the context in advance. The aim of Lippi and Torroni (2016) is, instead, to detect claims and evidence without knowing the context (or topic) in advance. AM systems architecture works by starting from unstructured documents, which are considered inputs, and aims to produce structured documents as outputs. The systems detect the several arguments in the unstructured corpora and explain their relationship through argument graphs. The process consists of two steps: the extrapolation and the classification of the arguments and, secondly, the detection of argument boundaries. The final step links the arguments detected in the previous stages, and concludes the process with the argument's graph.

2.3. Theory of Acceptance

The Technology Acceptance Model (TAM) aims to explain the factors contributing to adopting technology. Davis (1989) proposed TAM to measure user behaviour of Information Technology. The objective was to have a simple model to consider the most relevant emotions that consume experiences. The author underlines that someone decides to use a service if it facilitates the work and calls this *perceived usefulness*. If users perceive the system as helpful, it might be too difficult to use. This concept is called *perceived ease of use*. These two variables are strongly correlated. Kulviwat et al. (2007) propose a Consumer Acceptance of Technology (CAT) to explain the consumers' adoption intentions and overcome the limitations of the Technology Acceptance Model (TAM).

The Unified Theory of Acceptance and Use of Technology (UTAUT) merges eight theoretical models. It consists of four components of behavioural intention: performance expectancy, effort expectancy, social influence, and facilitating conditions (Venkatesh, 2003). A second version of the UTAUT includes three components: hedonic motivation, price value and habit (Venkatesh, 2012).

A literature review of the unified theory of acceptance (Williams et al., 2015) highlights the main shortcomings of existing studies. Among other things, most studies have a limited sample size.

One of the main issues in data sharing is the lack of trust among data users and the need to ensure privacy. Blockchain can be the solution to these challenges, and this is the reason why several works propose data-sharing platforms based on this technology. Shrestha and Vassileva propose a framework to evaluate a blockchain prototype for research data sharing. Their work extends the Technology Acceptance Model (TAM) and theorises that behavioural intention predicts actual behaviour.

2.4. Business Model

In recent years, the development of new markets and the rapid increase of innovations and technologies in terms of accessibility have led companies to differentiate their business models. This concept has assumed ever greater importance in research and support practices for managerial decisions, constituting a key factor for understanding the evolution of business choices and the relationships between the production factors. However, the understanding of this concept still appears elusive, and only in the last twenty years, the literature has focused on finding a shared definition of business models. At the root of this heterogeneity in the conceptualization of business models, there is not only the progressive push of research into the study of how organizations are structured and create added value but also the differentiation of business systems, which have become gradually more complex and heterogeneous.

2.4.1. Definition of Business Model

A business model is a set of structure, content and governance of transactions aimed at obtaining an added value deriving from the exploitation of business activities (Amit and Zott, 2001). Therefore, a Business Model (BM) constitutes a structure to make business choices. A BM generates added value and revenues to support the organization's activities (Rappa, 2004). The evolution of a BM assumes the relationship between the organization and its ability to generate profits, creating value for customers willing to pay for the goods and services offered (Teece, 2010). Consequently, a BM constitutes the logic through which a company sustains itself, meets its objectives and generates added value in the form of profits.

This logic embraces the corporate life system as a whole, defining, from another point of view, a BM as a set of relationships established between the technical inputs - the goods and services, and the economic outputs represented by the value of the business, profit and price. In this conceptual framework, a BM acts as a blueprint for corporate choices, influencing the relationships structured in terms of input-output about goods and services and into vertical and horizontal roles and relationships between economic actors.

The vast set of definitions that refer to the concept of BM highlights the existence of structured relationships between production factors aimed at achieving objectives and creating added value deriving from the activities and choices of management. This concept can be analysed from different points of view, as the structure of the BM embraces diverse areas. The scheme of corporate relations and procedures defines an objective aspect of the BM but also allows an understanding of how these relations intersect within a structured framework, implying the subjective dimension of the BM.

The conceptualization of the BM implies that the existence of a sequential relationship between managerial choices and their consequences determines the added value. In this sense, corporate

Funded by the European Union

decisions are part of a broader business strategy, which must be distinguished from the BM and helps to define the objectives and directions. The BM is the frame, explaining all the relationships structured as a result of the business strategy and which contribute to capturing the added value generated by company activities, goods and services produced, and knowledge employed.

The BM represents a logic that underlies an integrated set of choices on a decision-making level. They influence the structure and the characteristics that define the theoretical BM adopted. In Brea-Solìs et al. (2015), the authors classify the choices of a retail discount into eight categories, called levers, based on the definition of Porter's (1985) value chain[1].

The first of these categories is the **pricing structure** that identifies the price choices and the level of discrimination adopted. The second category is the **pressure on suppliers** which defines the pressure exerted on the suppliers to obtain favourable contractual terms or to adopt partnerships capable of increasing the added value.

The third category is the **technologies** used based on strategic choices in daily production processes. The relationships established based on the vertical and horizontal interactions between the operators (employees, suppliers and the subjects who collaborate in the production processes) define the technologies introduced at the operational level. The company adopts various policies that regulate these relationships, for example, through incentives and bonuses to set up the fourth category: **human resources practices** (fourth category).

The choices related to the location of the offices on a geographical level (rural, urban and semi-peripheral areas) influence the company's territorial expansion possibilities and are called **expansion policies** (fifth category).

The sixth category, **product selection,** defines the strategic choices associated with the mix of products and services and their differentiation in terms of type, quality, prices, categorization, etc.

The seventh category is associated with the choices to minimize costs, based on which the company structures its production cycles.

Finally, the last category is the set of relationships with customers and is called **customer service**. Specifically, this category denotes the set of company policies that establish what type of relationships to maintain with customers, such as return policies, consultancy and customer support services, and the language used in relationships.

In this sense, the BMs have a dual nature which consists of the possible choices for each type of business. The set of possible categories is called "levers". The BM is associated with the scale model that represents the theoretical description of the organization and its relationships. In other words, a scale model represents a small representation of the system of choices, or rather the structure which will subsequently constitute the model. A scale model has the same function as a smaller model. Its objective is to reproduce the BM. The realization of a scale model, or the case study, represents the

[1] Porter (1985) defines a value chain as the sequence of activities carried out by a company in a specific sector. In this sense, the company acquires an added value determined by the sum of the activities carried out by the company in the production cycle.

role model, which completes the dual nature of the BM concept since, starting from the theoretical model, it defines the set of concrete choices made on the decision-making level, starting from the classification of each lever.

The design theory defines the theoretical model. The design principles constitute a framework of fundamental prescriptions of techniques, methodologies and operating principles which define the subsequent phases of the design and implementation of a BM.

2.4.2. Evaluation of the Business Models

Based on the definitions presented in the previous section, a BM differs in the design principles, structure and relationships determined based on the BM. The design of a BM allows setting up strategic choices thanks to the possibility of comparing the competitiveness of different BMs on the market, implementing one's own BM, and defining the objectives and ambitions related to the BM adopted.

When evaluating a BM, a tool is needed, which is a reading key that allows the analysis of different BMs from multiple points of view.

The development of evaluation techniques that support the creation of value of a BM is needed to understand how the businesses and the related new technologies, practices and IT tools will evolve. A fundamental element of the evaluation, following the development of the necessary design principles, is linked to the development of theoretical frameworks that can serve as case studies to compare the theoretical models capable of collecting the characteristics of the BMs.

The model proposed by El Sawy and Pereira (2013), called the "VISOR business model", distinguishes the relevant aspects of BMs into five fundamental categories of design principles. The first of these is called **value proposition** and refers to the set of choices that aim to link the niche segment of customers to the products and services offered or the reasons that can push customers to pay the value company added. The second element, called **interface,** concerns the set of ways in which customers interact with the organization and, therefore, how it expresses the added value. The **service platforms** factor refers to the set of processes and relationships necessary to create and distribute the added value determined by the output of production processes. The fourth category, the **organizing model**, refers to business processes and external and internal relationships to create the value chain. Finally, the category **revenue and cost-sharing** refer to the ability of BM to ensure that revenues manage to exceed costs to the extent that the business is profitable.

Similarly, one of the most popular models for creating and classifying BMs is the Business Model Canvas (BMC). Based on the definition reported by Osterwalder et al. (2010), a well-structured BM can be described based on nine factors, which determine the characteristics of the underlying structure. The **Customer Segment** is the subjects, individuals and organizations that will benefit from the final products and services. The second set is the **Value Proposition**, which represents the set of products and services to satisfy the demand. This factor contributes to the added value since the company, through its products and services, determines the needs of the customers who pay the company, generating the profits necessary for its livelihood.

Funded by the European Union

The third element is the **Channels** through which the company communicates with customer segments to satisfy the demand. In this conceptual framework, the methods of communication with customers and distribution of products/services are the interface of the company or the systems through which the latter sells its final products. The **Relationships** with customers are the fourth element and are distinguished from the channels, defining the techniques through which the company approaches the different customer segments. The **Key Resources** represent the assets necessary to ensure company activities and communication with customer segments and, more generally, generate revenues. The sixth element is the **Revenue Streams** which are the resources generated by marketing and sales. The **Key Activities** represent the seventh category and refer to the actions carried out by the company to ensure the value proposition. These actions foster relationships with customers and guarantee the company's market. The **Key Partnerships** are the networks created by the company to optimize business operations and the allocation of resources and achieve economies of scale by reducing risks. This element is necessary to ensure the company's competitiveness in the markets. The construction of relationships with other subjects contributes to the company's growth. Finally, the cost structure keeps track of expenses, identifying costs associated with company operations such as, for example, acquiring resources and marketing products.

The models presented are an overview of the structural characteristics of companies' BMs in the digital goods and services sector. Based on these definitions, BMs can be classified and compared.

2.4.3. Typologies of Business Models

The design principles break down a BM into different categories, within which it is possible to make strategic choices and evaluation processes, dividing it into the elements that together make up the overall structure. Griessmann and Legner (2016) analyse the development of companies that provide **digital Platforms as a Service (PaaS)** concerning the BM adopted, trying to draw a conceptual framework for their implementation and evaluation. PaaS refers to cloud platforms to connect customers with sellers using one or more digital services and tools. The solutions adopted by companies in this field are many and can be traced back to three main spheres. The supply of **Software as a Service (SaaS)** has, in the first case, extended the services to cloud platforms. In the second case, software vendors have developed their cloud solutions to integrate with their solutions. Finally, companies may not produce cloud solutions but acquire other PaaS stat-ups. The authors have identified six design principles that constitute the conceptual framework to analyse and compare the different business choices used by companies regarding PaaS models. The first principle identifies five types of customers of PaaS providers: individual developers who are not active in the development of commercial solutions, independent software vendors (ISVs) asking for components and applications (C&A), systems integrators (SIs) who mostly do consultancy and act as intermediaries for other clients, companies that use PaaS solutions as their private clouds and develop their own C&A and the final consumers of C&A.

The second design principle is the Value Proposition which can vary according to four types of platforms. The first form involves providing an environment for developers to develop and test their applications. In this case, the Value Proposition of the BM can materialize in the possibility of facilitating the development of C&A. The second form of the platform allows ISVs to develop extensions of their core software solution. The Value Proposition linked to this solution may concern the digital tools and services integration developed into existing SaaS solutions. The third possible form is the ability to connect the cloud platform to multiple applications. The Value Proposition, therefore, concerns the integration of on-premise and on-demand tools and applications. Finally, the fourth type of platform consists of online communities and provides environments such as markets where it is possible to share, sell and buy other products and services. In this case, the Value Proposition lies in providing a distribution channel for goods and services of others.

The third design principle concerns the development of C&A. According to Griessman and Legner (2016), the platform needs to offer as many innovative support complements as possible. The components and applications constitute the platform's content and define its value for customers. Therefore, the more varied the service offered in terms of applications, the higher the platform's value.

The fourth principle of design concerns the need to constantly monitor the needs of its existing customers to maintain an active customer basis and strengthen the relationships between the company and customers to create a relationship of trust and reach critical mass. The fifth design principle refers to the definition of precise rules to regulate the relationships between the company and external subjects regarding property rights and collaboration relationships. Finally, the sixth design principle deals with the effective internal governance structure. In this sense, this structure must avoid conflicts of interest, develop effective forms of communication between agents and, more generally, strengthen the company's credibility within and outside. There are five fundamental elements: aggregate groups with similar objectives in a single business area, encourage organizational culture, improve internal communication, protect and develop new PaaS solutions and strengthen its internal use.

Gordijn and Akkermans (2001) compare two types of models, called the Terminating model and Originating model, for evaluating different BMs in the e-commerce field. The authors consider many types of businesses involved in the commerce of digital goods and services, such as online newspapers, telephones, connection fees etc. In the Terminating BM, customers pay on a bundle basis. The value interface shows that users purchase two or more types of goods, for example: telephones and connection fees, in combination. These goods and services are complementary and drastically decrease their exchange value if taken individually. The "Last Mile" local operator influences the price structure. The operator generates the secondary flows. The Last Mile plays a fundamental role since it is responsible for most of the relationships with the customer and is at the same time responsible for local data traffic. The Last Mile is an intermediary between the customer and the company, managing the payments. In the case of the Originating model, the role played by the operator, i.e. the one providing the goods and services assumes greater importance. The customer interacts directly with the latter, while the Last Miles do not have direct relationships with customers. Unlike the Terminating model, in which the Last Mile operator acts as an intermediary between the customers and the operator, managing prices, payments and data traffic, in the Originating model, the operator controls the flow of money through payments.

3. Evaluation of the data exchange culture

3.1. Methodology

Funded by the European Union

We propose a methodology to identify the claims and the main topics in a set of interviews to understand the data-sharing culture in the transport and mobility sector. We define a protocol for the interview with macro-categories of questions. These questions are identified thanks to the analysis of the literature and based on the goal of the work. Next, it is necessary to identify a heterogeneous sample of interviewees based on some macro-categories of actors involved in the transportation sector. The interviewee, contacted via email, had a one-hour interview. We record the interview and transcribe it for textual analysis.

Our approach consists of the combination of two Natural Language Processing (NLP) techniques: namely, argumentation mining and topic modelling. First, we extract claims and evidence from the input text using the MARGOT (Lippi & Torroni, 2016) off-the-shelf argument mining tool. Second, all the argument portions detected in the first stage are processed via topic modelling to identify relevant topics. The key idea behind this approach is that argument mining can filter out irrelevant content and keep only those statements (i.e., claims) that are the most interesting to analyse.

MARGOT is an online server that extracts argument portions (claims and evidence) from any textual document that is fed as input. Although the underlying model was trained on Wikipedia articles, it has shown good performance across a variety of different genres, topics, and domains (Lippi & Torroni, 2016; Lippi et al. 2022). Argument component detection is performed by MARGOT using tree kernels (Moschitti, 2006), that exploit the structure of constituency trees to look for similarities between sentences. The classifier computes two distinct scores for each sentence, that can be interpreted as the confidence assigned by MARGOT to the fact that such sentence contains a claim or, respectively, an evidence. By default, a sentence is predicted to contain a claim (respectively, evidence) if the corresponding score is positive.

3.2. Description of the Dataset

This section reports the preliminary results of the topic modelling analysis and argument mining performed on ten interviews. The interviews took place between June and September 2022. The plan is to execute at least forty interviews with the different actors of the data-sharing culture by the end of the MobiDataLab project. D3.1 reports the description of the actors. Table 4 shows the number of actors who will receive the interview by the end of the project. The first column shows the original category of actors identified in D3.1, the second column reports the new macro-categories, the third column indicates the number of interviews, and the last column shows the number of interviews considered in D5.2. The difference between the original categories of actors and the new macro-categories is minimal. For instance, Ambulances and Police are in the same category named “Emergency services”. The first three interviews lead to a revision of the protocol compared to D3.5. The new protocol is reported in Annex I.

The dataset consists of interviews performed with the categories of actors reported in the last column of Table 4. We did not perform all interviews because the Task is still ongoing and we plan to perform all 40 interviews by the end of the project duration. The final results of the analysis will be presented in D5.3 Analysis and conclusions on the data exchange culture. The number of people to interview for each category of actors was selected according to the number of sub categories identified in D3.1 Actors’ needs and cooperation framework report. Each interview lasted about one hour. The recording's transcription is the dataset deployed for the analysis. The interviews' objective is to

Funded by the European Union

identify the main themes or topics related to the data-sharing culture and its impacts on society and the environment.

Table 4: Categories of actors and number of interviews (September 2022)

Categories of actors from D3.1	New categories of actors	Number of people who need to be interviewed for each category	Interviews done between June and September 2022
Research centre/ Universities	Research centre/ Universities	2	2
Citizens	Citizens	2	-
Commuters/Passengers			
Tourists			
Tourist Agencies	Tourist agents	1	-
Hoteliers			
Tourism Associations			
Climate change NGO	Climate change NGO	1	-
Regional policy makers	Policy makers	3	1
European Policy Makers			
Municipalities/associations of municipalities	Municipalities/associations of municipalities	3	
Government / ministries	Government / ministries	1	-
Government Transportation Agencies			
Public Transport Authority	Public Transport Authority	3	-
Rail Infrastructure Authorities	Rail Infrastructure Authorities	1	-
Traffic management centre	Traffic management centre	2	-
Trade Association	Trade Association	1	-
Logistics operators	Logistics operators	3	-
Public Logistics operator			
Ride-sharing companies	Public Transport Operators / Transport Agencies	4	1

Funded by the European Union

Categories of actors from D3.1	New categories of actors	Number of people who need to be interviewed for each category	Interviews done between June and September 2022
Micro-mobility operators			
Public Transport Operators			
Transport Agencies			
Airlines	Airlines	1	-
Pricing/payment platform	Intelligent Transport Systems (ITS) providers	4	3
Navigation services providers			
Trip Planners			
Software providers	Software providers	1	1
Search Engines	Search Engines	1	-
Mobile phone operators	Mobile phone operators	1	-
Cloud Providers	Cloud Providers	1	-
Satellite Operators	Satellite Operators	1	-
Autonomous vehicles manufacturers	OEM/Car Manufacturers/Association of car manufacturers	2	1
Fire Service	Emergency services	1	-
Ambulances			
Police			
	Total	40	11

The interviews help to identify the main themes or topics related to the data-sharing culture and its impacts on society and the environment. To achieve this objective, we deployed the web server for argumentation mining named Margot (margot.disi.unibo.it). The tool allows elaborating corpora of limited dimensions to identify claims and evidences. For this reason, the analysis considers one interview at a time. The results are sentence claims and evidence scores. Since not all the identified claims were relevant or correctly identified, we cleaned the dataset and eliminated irrelevant claims such as direct answers like “yes, sure”.

We identified the topics of the interviews thanks to the topic modelling analysis. The text deployed to identify the topics is the claims extracted thanks to the argumentation mining.

The first step of the topic modelling analysis is to pre-process the text so that there is no punctuation and stop words. Next, we created the bigrams and trigrams models to identify couples and triplets of words. The lemmatisation allowed the transforming of verbs into their infinite form and the plural nouns into singular. Thanks to LDA, it has been possible to identify eight topics named and validated by the claims.

3.3. Experimental Results

The pipeline uses spaCy (Honnibal & Motani, 2017) and Gensim (Rehurek & Sojka, 2011). We tokenized words, assigned tags to them and implemented the entity recognizer to classify tokens according to the transition-based algorithm (Covington, 2001). We lemmatized words and added one rule. Specifically, we avoided getting the singular data (datum).

We tuned the thresholds to determine the bigrams and trigrams with Gensim. Gensim library allows to detection of phrases based on (Mikolov et al., 2013), an extension of the Skip-gram model that creates a representation of words with which it is possible to predict another word in a phrase or document which is coherent with the context. The phrase detector (Bouma, 2009) is a normalised mutual information index to connect words. Two words are connected if they appear together and the probability to see one word is equal to the chance to see the other word. Finally, we created a dictionary of data and a corpus with terms frequencies to implement LDA.

We manually performed several experiments and checked the coherence indicator and the frequency of words on each topic. According to the frequency of words occurring in each topic, we removed the ones typical of colloquial languages such as *maybe*, *example* and *super*. We also considered if the same word appeared in different topics. The experiments consisted of manual tuning of the alpha parameter - finally set to 0.1, and the number of topics.

Using topic modelling, we obtained six topics from the claims identified by MARGOT. Table 5 reports a brief description along with some examples of claims associated with them.

Table 5: Topics

Topic name	Topic definition	Representative citation
Topic I & II: Prerequisites for data sharing	The existence of previous collaboration is important for [data sharing] although there are always risks and opportunities.	"For the matter of trust, I think it's well it's always important to have some previous collaboration beforehand because you 're never sure when you collaborate." "I think that the balance of risks and opportunities remains an individual decision."
	Data need to be standardised to have	"I would say that data and not only in terms of a journey planning or in transport, but in a more generic way, publishing

Topic name	Topic definition	Representative citation
	quality and be useful for travellers to move in an efficient way.	data on the web you need to have a quality data which means up to date, which is a standardised meaning it 's accessible through given standards and that you can quantify." “The quality should be good enough to compute very efficient journeys.”
Topic III: Social impact of data sharing	By demonstrating that the use of public transport is more efficient in terms of travel time and business, it is possible to have an impact on the society and on the environment.	"For social impact, I think that to be also used, you have to demonstrate that public transport is more efficient, more time saving, more environmental saving than your - take your car for example - or to take the bus, there are 20 cars that have to be parked in a place that has this dimension like this public bus is this dimension - you have to demonstrate time-saving, money-saving, and environmental saving I think is the most difficult thing to do." “The journey planner allows for the creation of new services for pre-trip planning and can create a positive impact on society and the environment.”
Topic IV: Challenges	Data fusion of different dataset is difficult especially in terms of providing good quality services	“And once again, in the ecosystem of car sharing of car personal car data, there has been a minimum level of quality that has been achieved that explains that a lot of transactions of data sharing monetized is observable, but on the contrary, when we try to bridge the gap between that type of data and other types of data, we see that it’s still very difficult because, in part of that problem of data quality and comparison, that is not easy between the two environments." “We observe is that indeed some sort of reluctance to share that kind of data, for instance for public transport ticketing to private actors for different reasons, for costs reasons, for competition reasons, so I think it really depends.”
Topic VI: Ecosystems of data sharing	Data can be produced by public authorities or private users.	"So for the observations that I had on data sharing, there is something that is quite clear is that according to the type of mobility data, there are more or less, let 's say there are two ecosystems that work pretty independently, one centred around car data or personal vehicle data, and the other centred around public transport data." “So we would see, we would say naturally that sharing is not let's say natural for private companies because it prevents them from collecting revenue out of the data, but sharing data for free with cities helps them to grow their business , and they could do so.”
Topic VI: Vision	The future of data sharing is the possibility to offer a mix of public and private transport modes.	“From the private and public perspective, I see that the goal a more sustainable mobility, and encourage people into changing their behaviour to opting for soft modes, active modes , and dropping their cars, so I think If I think us the competition is not public transport or it is private cars ownership because it takes an enormous amount of space , most of the time they 're parked , and that is why we're also working on electrifying the platform and encourage people to consider in a mix of mobility kind of options, that would slowly push out the private ownership of the cars to the for people who would have absolutely any kind of I do not know."

Topic name	Topic definition	Representative citation
		"We 're in the situation in which companies are pressured to share data as it's an emergency and time is passing, but then it's very hard to justify also to our users the location data that arrives in the hands of public bodies, that would reveal the travel patterns over times from point A to point B. I think that , from a privacy, perspective is very concerning because we 're arriving in a situation in which, as an emergency, we actually give a lot of to public bodies to know everything about the movements of people, I think that in a democratic society it's very worrying and we 're lucky to have data control, protection authorities , and we have to be sure that any kind of data that has to be shared has a purpose and controls and there is transparency about any information about data 's going to be used ."

From the analysis presented here, Topic I and Topic II have the common characteristic of having a data-sharing cooperation framework: it is clear that previous collaboration is a necessary prerequisite for data sharing. For this reason, *Topic I* and *Topic II* are *called Prerequisites for data sharing*.

As highlighted in the claims of Topic I, trust in people and companies with which they share data is a fundamental component. The situation is further improved when a cost-benefit-risk analysis can anticipate every data decision.

If *Topic I* deals with emphasizing the importance of having prerequisites, Topic II suggests what these requirements are, highlighting the need for standardized and high-quality data.

Du et al. (2012) report an analysis of existing collaboration as a prerequisite for data sharing. Among the factors that affect the willingness to share, the authors consider the partnership extent in the logistics sector that affects the degree of sharing and the frequency of change in the value of data sharing. They found that partnership is needed but not sufficient for proactive information sharing. There is a need to design contracts and define the procedures for sharing information. Similarly, Moschovou et al. (2019) identify the barriers and enablers of a public-private partnership (PPP) for data sharing in the freight transport sector. The combination of different data sources is an enabler of data-sharing PPP. However, the authors point out that the creation of PPP should follow a set of guidelines concerning legislation and cooperation. Overall, public-private collaboration requires data governance and allows addressing the barriers to data sharing (Eckartz et al., 2014).

Bauer et al. (2019) focus on the relationship between trust and cooperative data-sharing behaviour. They found that privacy is a significant predictor of data sharing.

The second subtopic is focused on data quality, underlying its importance in information sharing. An analysis of information sharing and information quality in the supply chain (Li & Lin, 2006) found that supply chain stakeholders' trust and their shared vision influence data quality. Trust and data quality are important in the case of Internet of Things (IoT) or Internet of Vehicles (IoV) data. In this case, blockchain is a promising technology to improve trust in data sharing and thus increase quality. For instance, Chen et al. (2019) propose a blockchain platform to exchange vehicle data and ensure that there is trusted and high-quality data. Rouhani et al. (2021) underline that transparency in data sharing increases trust in users who provide data. Therefore, the authors propose a blockchain framework to evaluate data quality based on reputation, endorsement and confidence. Kang et al. (2018) blockchain reputation-based systems allow sharing of quality data. The system prevents data sharing, if not authorized, and manages the vehicle's reputation.

Topic III (*Social impact of data sharing*) highlights the importance of the social impact and the complexity of society which requires different perspectives and points of view. The common ground of our research is the European project MobiDataLab (Labs for prototyping future Mobility Data sharing cloud solutions). Figueiredo (2017) highlighted the importance of focusing on social impact of data sharing and the potentiality of data sharing for society. This topic highlights the importance of implementing actions to initiate a social impact chain (such as encouraging the use of public transport). It is necessary to promote social changes in terms of technological innovations, habits and cultural approach to change, as underlined by Witkowski (2017), Carmichael (2015) and Simanis & Hart (2011). Intrinsically, the interviewees represent the needs in a complex framework of challenges, risks, and barriers. On the other hand, the increasing use of digital transport systems allows for obtaining disaggregated data which increases risk of violating users' privacy. In this context, policymakers should protect vulnerable travellers.

Topic IV (*Challenges*) highlights the numerous difficulties that lie behind data sharing, both at the level of private companies and at the level of public administrations and cities. According to Gellerman (2016), the main challenges of data sharing in the transport sector are the existence of multiple schemes of ownership and legal framework, the description of the metadata is not always accurate and data format should be defined in advance, before data collection.

Challenges of data sharing can be related to ensure quality and integrate different dataset. In the context of smart cities, there are many applications of data fusion techniques. An example of complex data fusion system in the transport sector is the autonomous drive. According to Lau (2019), the main challenges of data fusion are: the quality of data, lack of data standardisation, privacy issues and the availability of data fusion techniques such as the approaches based on machine learning.

In the complexity highlighted by *Topic V* (*Ecosystems of data sharing*), clear differences emerge between different ecosystems for data sharing for which a first distinction can be made between the public and private organizations. Not only the way of collecting and sharing data is important, but also the objectives that lie behind data-sharing, as well as the mission and vision of the organization greatly influence the approach to sharing. If for public institutions it is natural to share data - or at least it is obvious to take into consideration the possibility of sharing data, for private companies more warranties are needed - and here the connection with the Topics I and II is evident. In this context, Du et al. (2012) points out that, when sharing data, organisations are more concerned with security issues, while users are interested to privacy. The existence of heterogeneous data-sharing ecosystems and the importance to take them into account to foster the development of application and services is main driver of data sharing platforms development (Munoz-Arcentales et al. 2019).

The *Topic VI* (*Vision*) is in line with works (Rouibah & Ould-Ali, 2007; Paulsson, 2018) which highlight the need for a vision that can encourage data sharing, knowing how to propose competitive behaviour for both public bodies and private entities. The necessity for an advanced data-sharing environment is a milestone for the International Maritime Organization (IMO). The IMO has formulated its vision adopted by an ever-increasing number of maritime states, regions and the EU (Lind et al., 2018). Furthermore, this vision allows understanding preferences, while imagining and evaluating different scenario perspectives (Keseru et al., 2021).

Funded by the European Union

4. Transport Cloud Acceptance

The acceptance of the Transport Cloud is assessed based Unified Theory of Acceptance and Use of Technology (UTAUT) constructs consisting of on an online survey sent to the category of actors already identified in D3.1. The online survey is reported in D3.5 and the description of the Transport Cloud to let the respondent understand was the Transport Cloud is, has been provided by the project coordinator AKKA. The main limitation of this survey is that there is not a prototype to assess by the participants. Furthermore, the respondents who have participated in the survey and are also partners of MobiDataLab were facilitated in understanding what the Transport Cloud should be. In our opinion, the survey is relevant even if the prototype does not exist yet because it is focused on the identification of the main constructs of the UTAUT that is performance expectancy, social influence and facilitating conditions that are correlated with intention to use the technology. The Transport Cloud without definitions of the product's specifics can be thought as a tool that allows data sharing. Therefore, the aim of the interview is to understand if there is willingness to use this type of technology by considering different categories of actors of the data sharing culture.

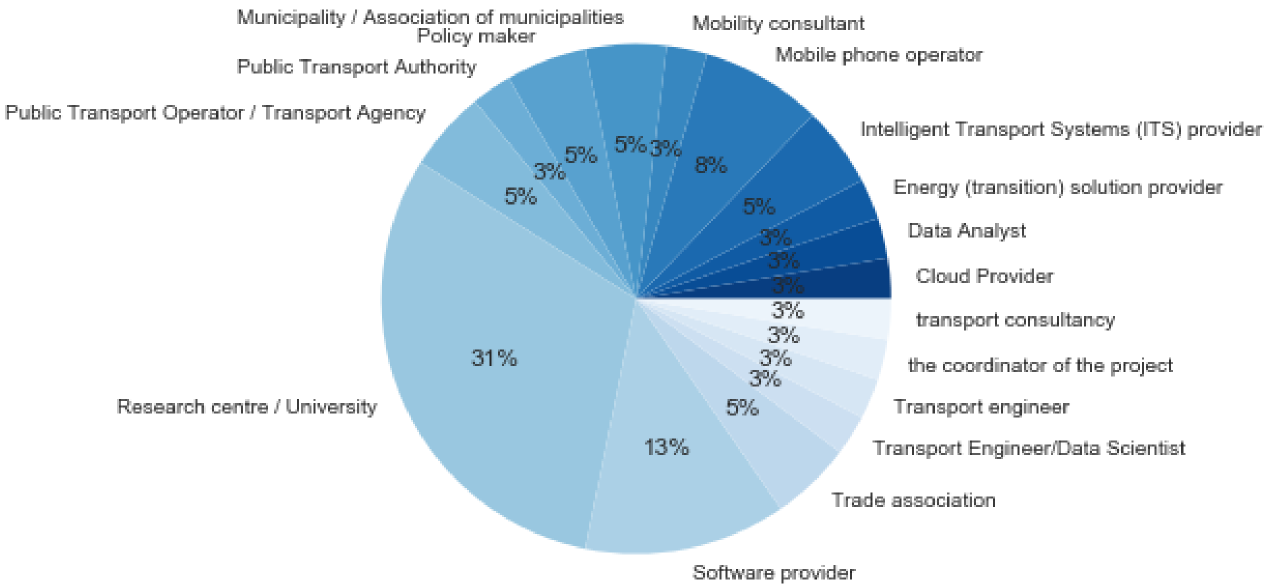

Figure 1: Participants in the survey on the Transport Cloud acceptance

The survey was filled in by 39 participants at the time of preparation of this deliverable (October 2022). Most of the respondents are Research centers / Universities (31% of the responses). Although the sample is heterogeneous. The majority of respondents (74%) are males and this reflects the gender unbalance that exists in the transport sector. The age of the majority of respondents (42%) is included between 36 and 45.

Funded by the European Union

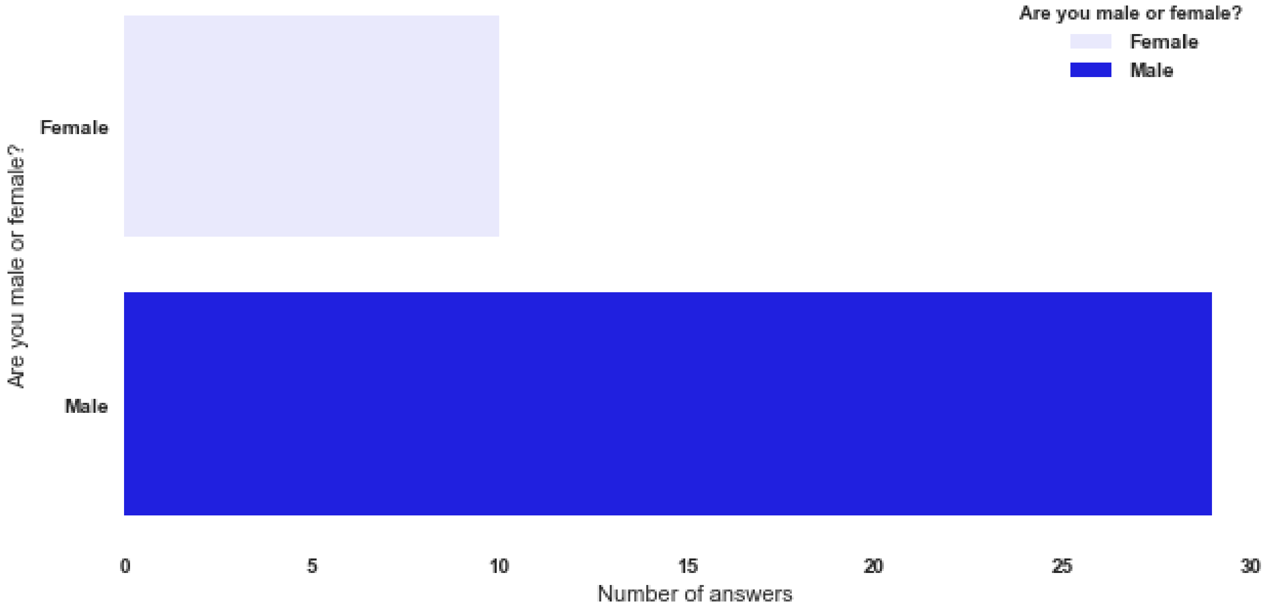

Figure 2: Gender of the respondents

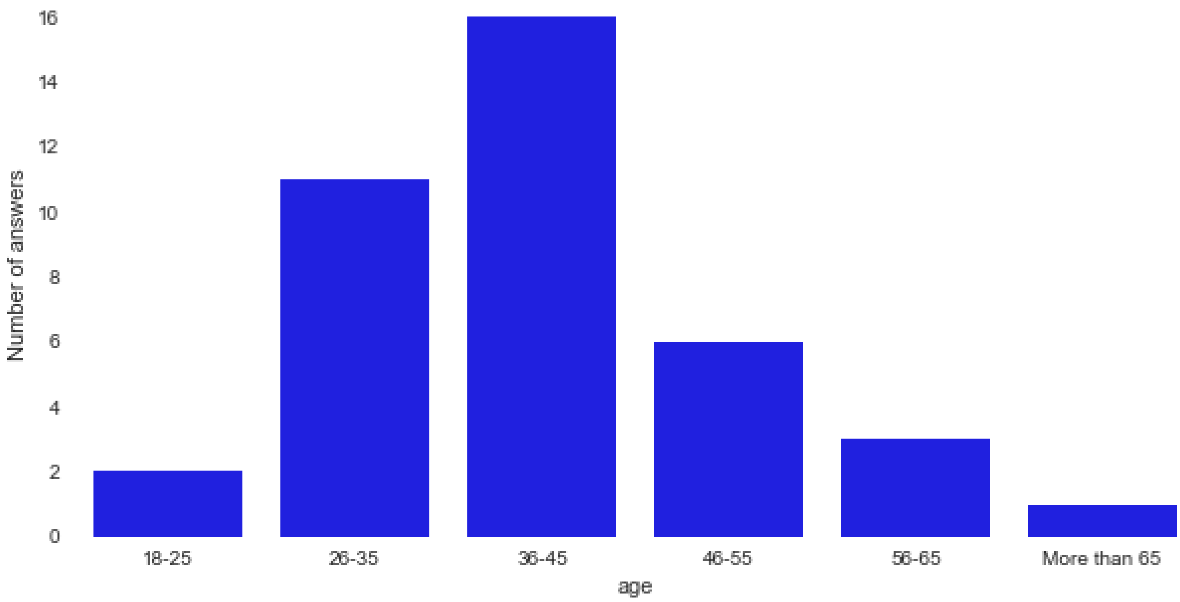

Figure 3: Age of the respondents

The first set group of questions aim to assess the **performance expectancy** and the majority of responses (about 70% on average) are assigned to 4 or 5.

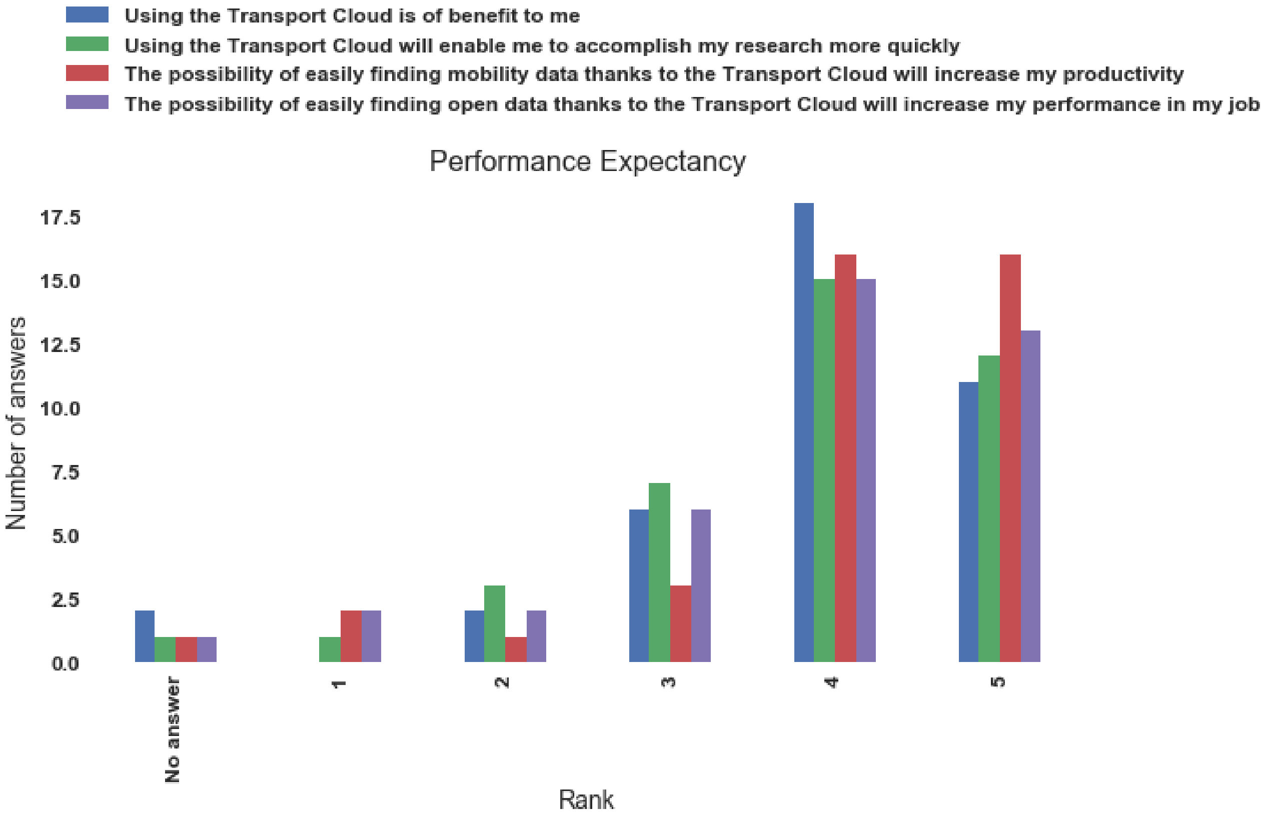

Figure 4: Performance expectancy

The second set of answers is about **effort expectancy**. In this case, the distribution centres around score 3, denoting a possible barrier to the intention to use the Transport Cloud. The description of the Transport Cloud provided at the beginning of the questionnaire could explain the lower-ranked scores of the questions related to effort expectancy. Two respondents wrote messages explaining that it was sometimes difficult to answer the survey because the product did not exist. We claim that the aim is to assess the intention to use a cloud data-sharing service for transport data which we refer to as the Transport Cloud.

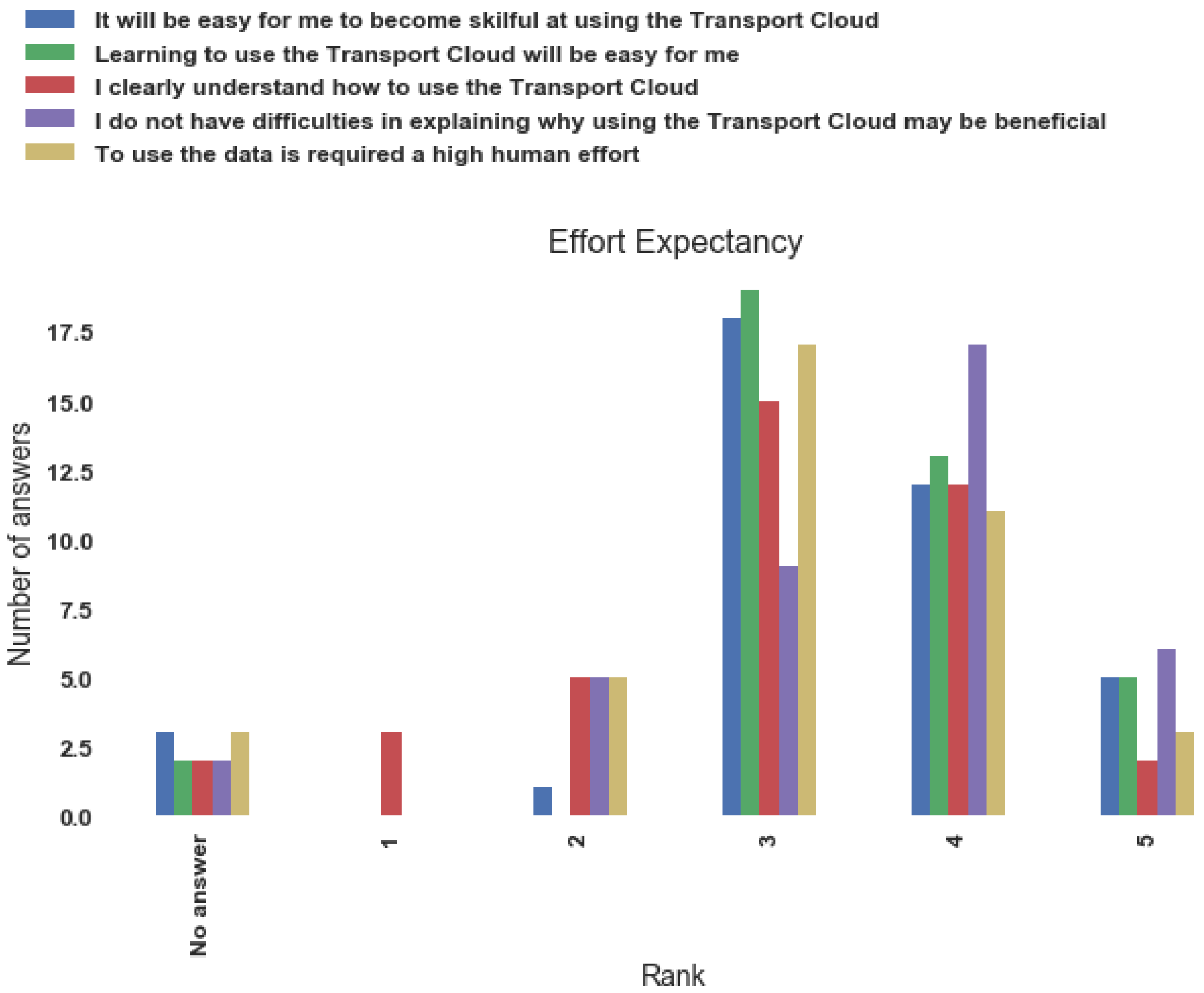

Figure 5: Effort expectancy

The third set of questions is related to the **facilitating conditions**. In this case, we want to assess the perception of users of the easiness of using a cloud data sharing Transport Cloud. Also, in this case, most of the answers are centred around the average score. Most importantly, the respondents believe that there will not be a specific person available for assistance or that the transport Cloud will hardly be compatible with other systems.

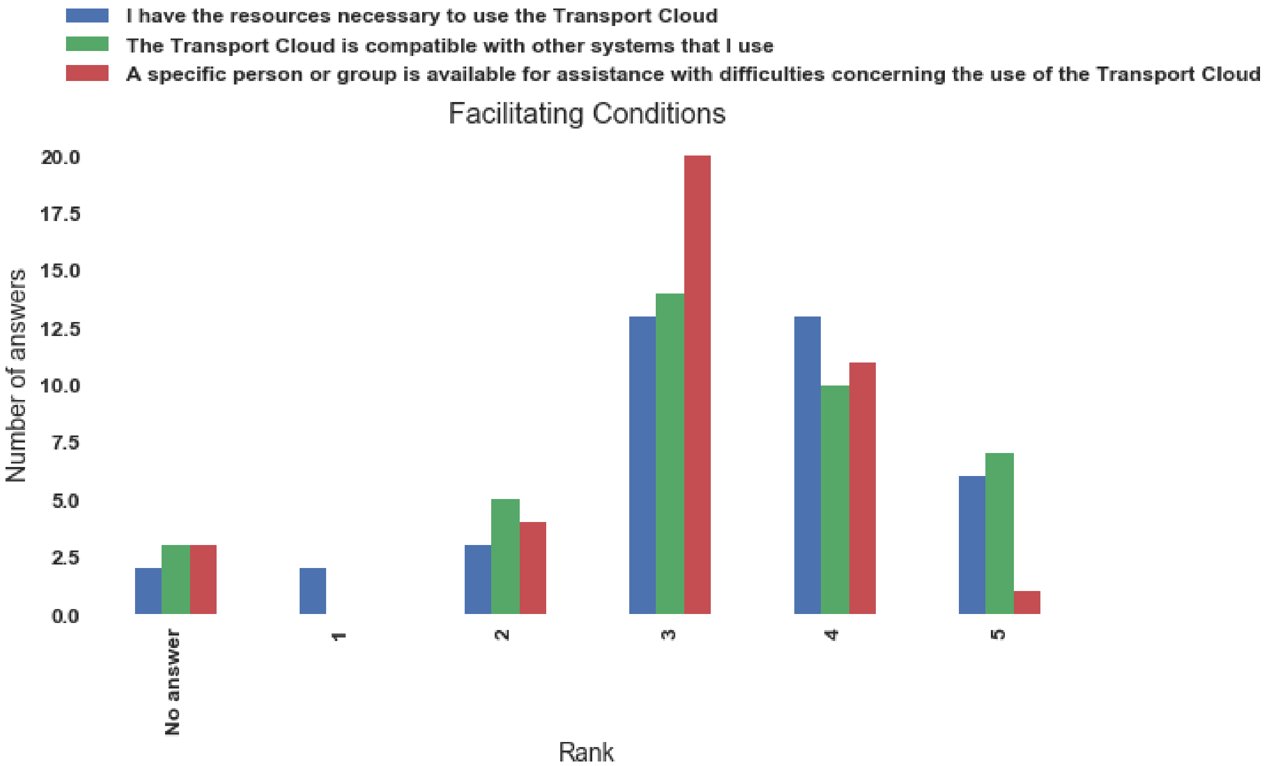

Figure 6: Facilitating conditions

The four sets of questions are related to **behavioural intention**. Although the concept of the Transport Cloud is not clear, the scores given by the respondents are promising and comprised between 4 and 5. Most of the participants in the survey believe that they will use the Transport Cloud in the future. The willingness to use means there is a need for this type of product. When considering the group of questions on facilitating conditions, it is clear that the data-sharing service should be easy to use and that there should be no need for assistance to explain how to use it or solves specific issues. Also, compatibility with the systems for data sharing should be considered.

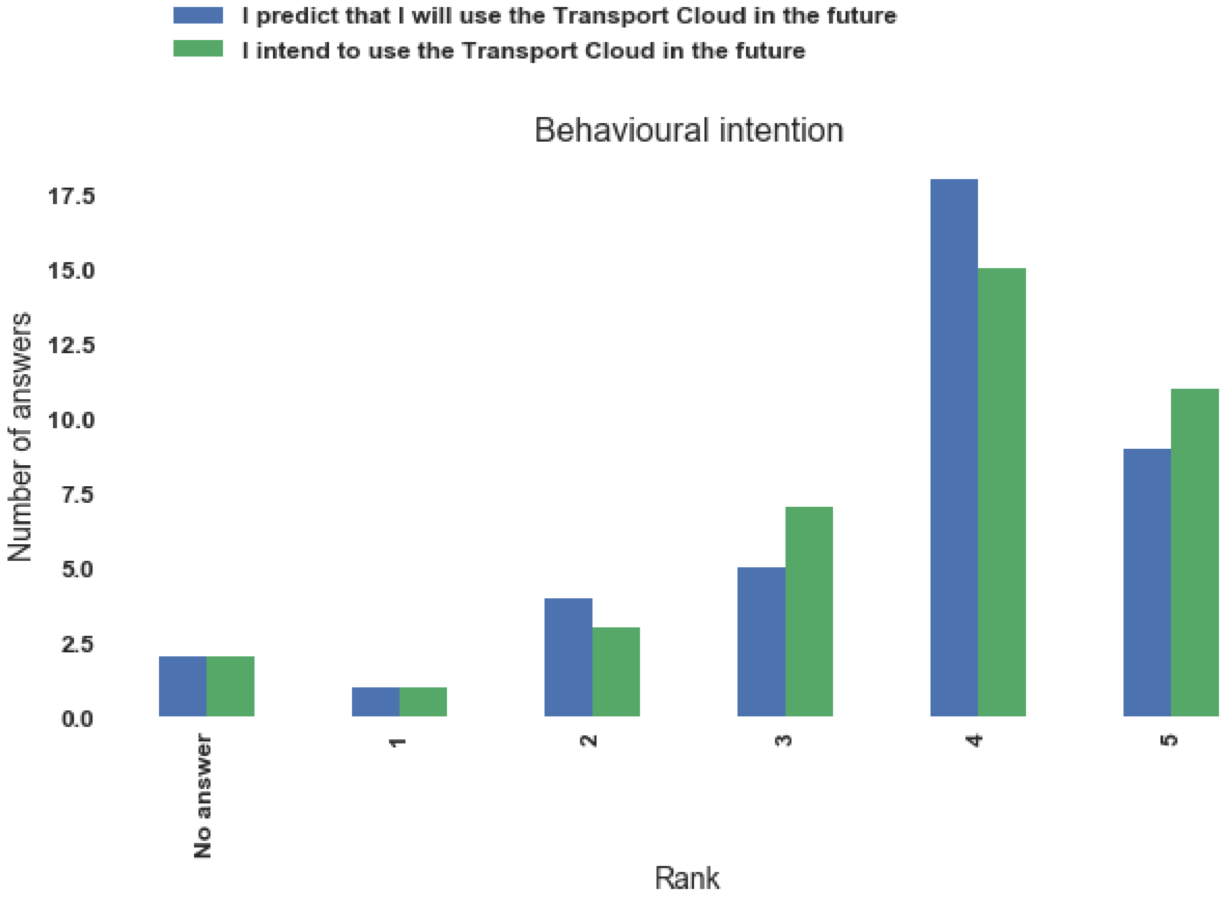

Figure 7: Behavioural intention

The voluntariness of use has a negative correlation with the Transport Cloud acceptance. If, for instance, the respondent works with transport data and must share them, there is no voluntariness in using a data-sharing platform. If there is the need to share data, the Transport Cloud could be a natural choice, guided by the existing situation. Most respondents claim they would deploy the Transport Cloud voluntarily, although most scores are 4, not the maximum. This result highlights that some respondents might use the Transport Cloud because they need it to perform their activities, especially if the Transport Cloud is something easy to use.

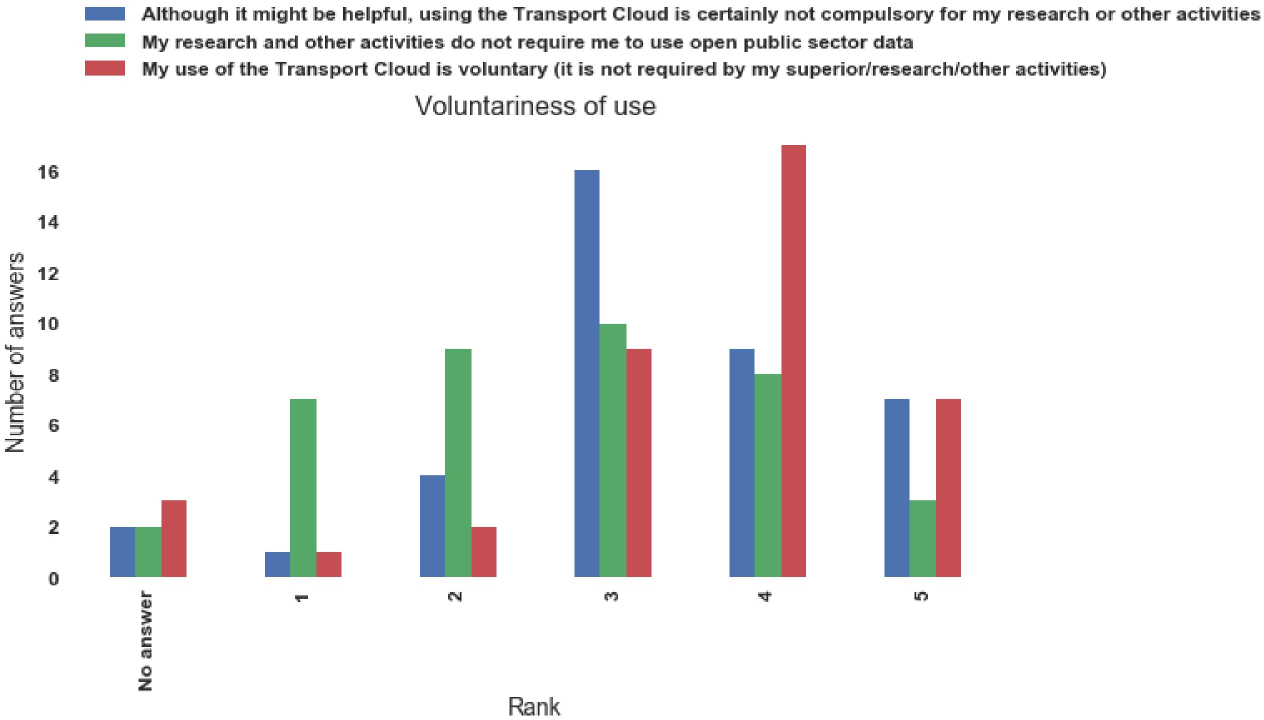

Figure 8: Voluntariness of use

Another set of questions concerned the functional requirements of the Transport Cloud. One of the aims of the Transport Cloud is to promote access and use of digital mobility services and data by vulnerable and excluded users (i.e. non-digital and people with reduced mobility). With this question, the respondents provide their opinion on expectations regarding how the Transport Cloud will function. Concerning digital inclusion, there is not enough confidence in the capability of the Transport Cloud to reach this objective. Four respondents do not agree that the service will have this capability. Regarding the response time, there are high expectations because the respondents think the Transport Cloud will provide the data without waiting too much. Similarly, data security, one issue of data sharing, is expected to be guaranteed.

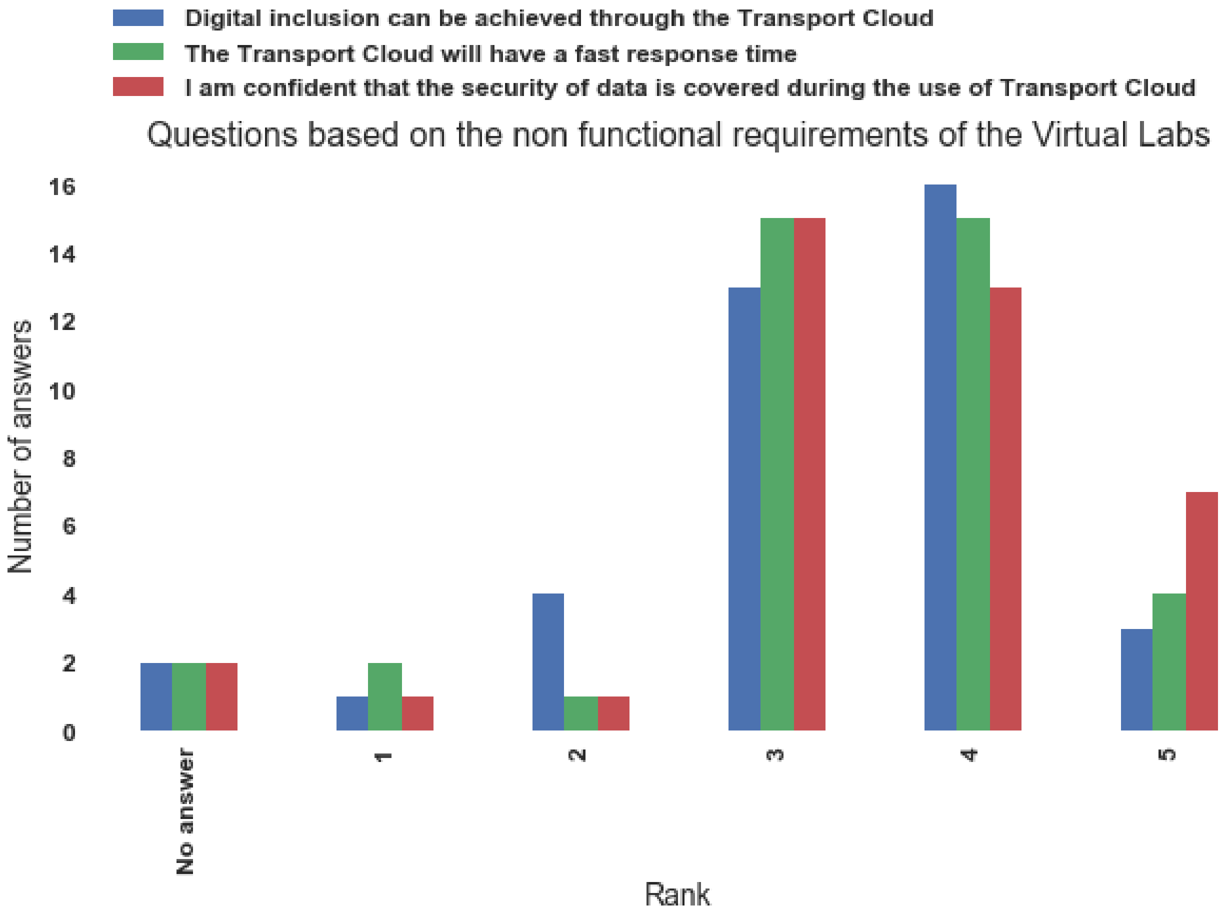

Figure 9: Questions based on the non-functional requirements of the Virtual Labs

Finally, the last set of questions are related to the Global Open Data Index which measures the findability of data and to understand how open data are and how much are useful for the public.

The data perceived as mostly needed by the respondents is transport data, followed by geographical data. Overall, the condition of mobility data is just fair. Few respondents think that the dataset is in good condition. Concerning the data update, the respondents answered that the dataset is not frequently updated, but 41% think that the dataset is. About 60% of the respondents claim that it is not possible to download the dataset at once. The 50% of respondents deploy free dataset, and 64% uses machine-readable data. About half of respondents (53%) report that there is no need to register to access the data. Finally, they believe that data is usually out-dated (about 53%). The figures reported on the following pages of this report show the number of answers for each question.

Funded by the European Union

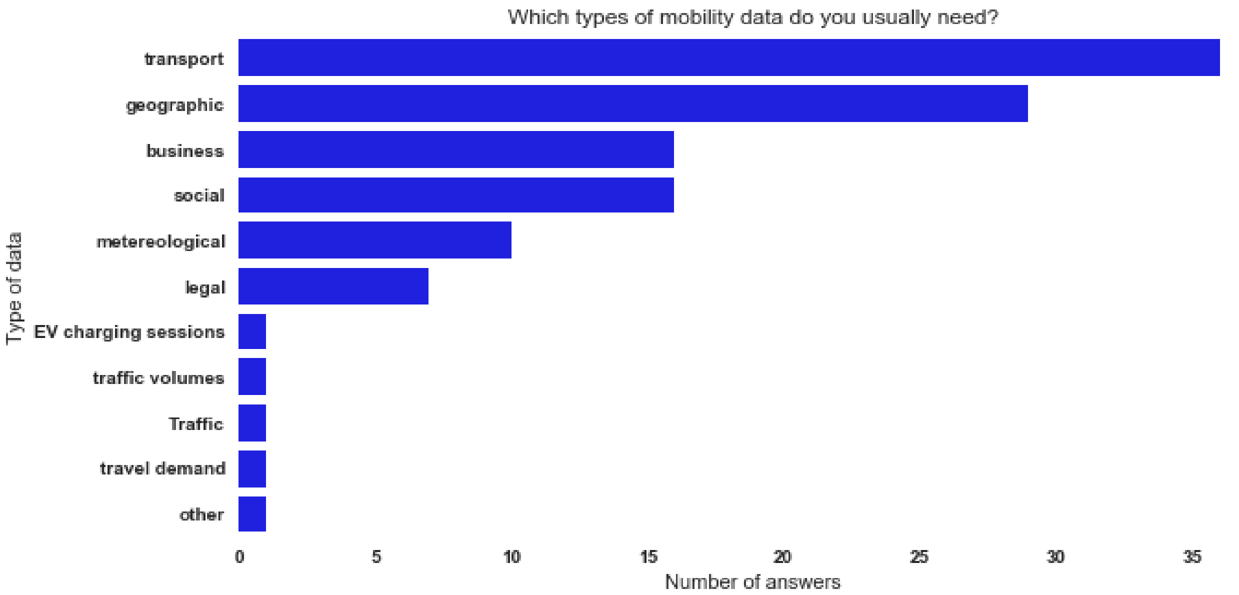

Figure 10. Which Types of mobility data do you usually need?

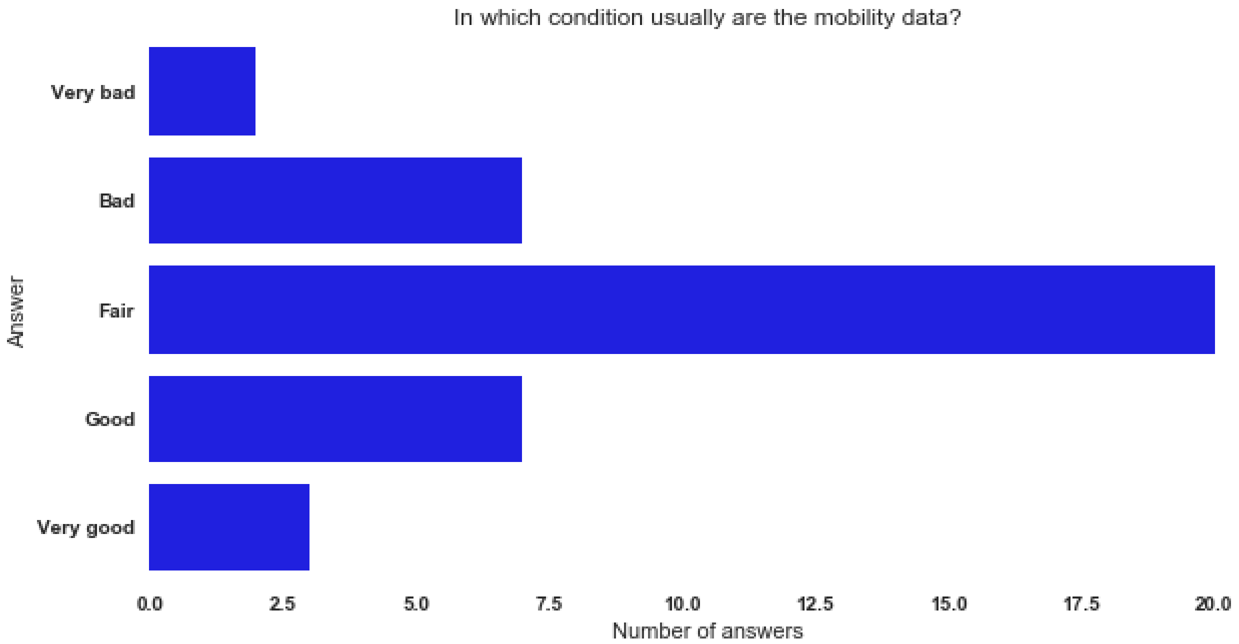

Figure 11: In which condition usually are the mobility data?

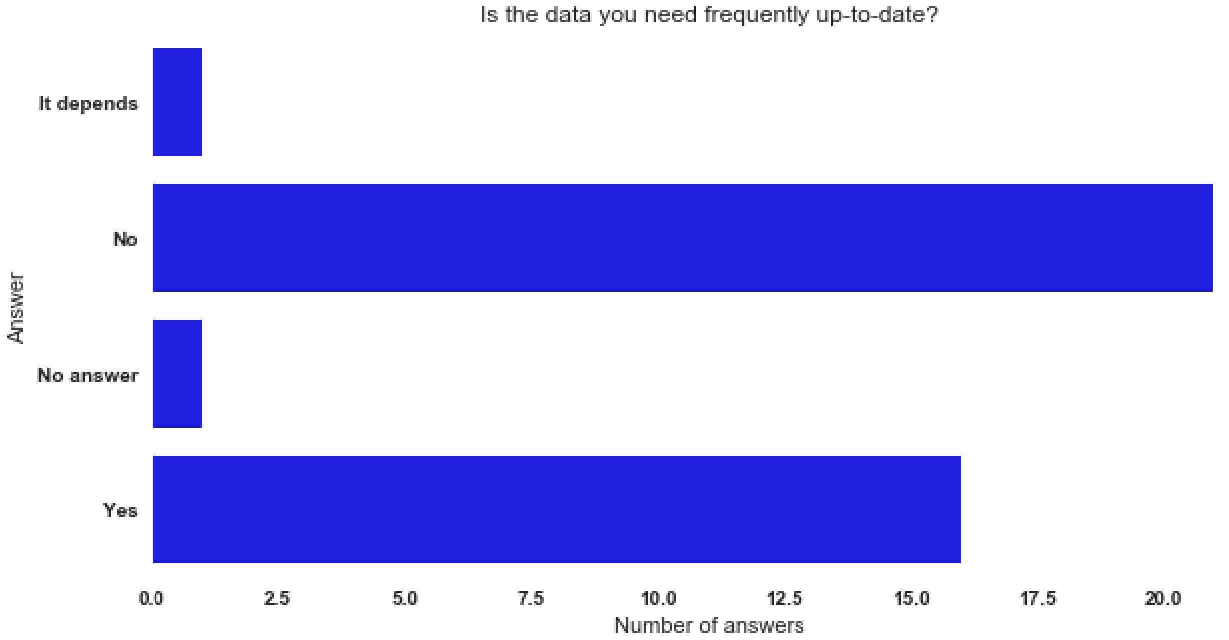

Figure 12: Is the data you need frequently up-to-date?

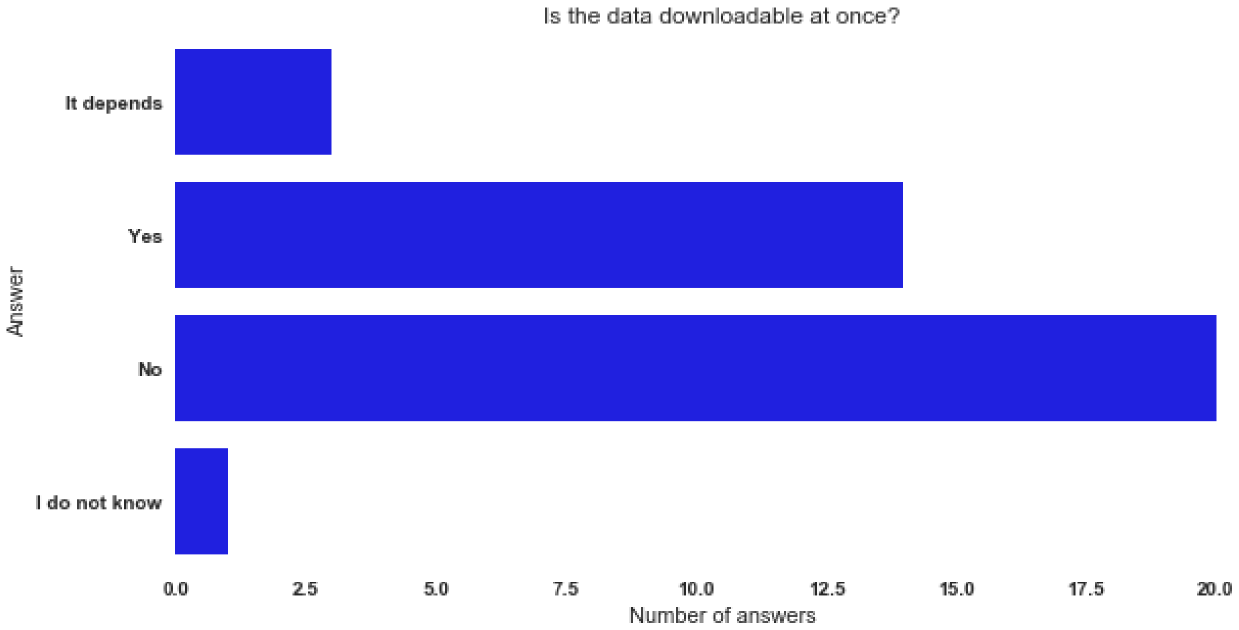

Figure 13: Is the data downloadable at once?

Funded by the
European Union

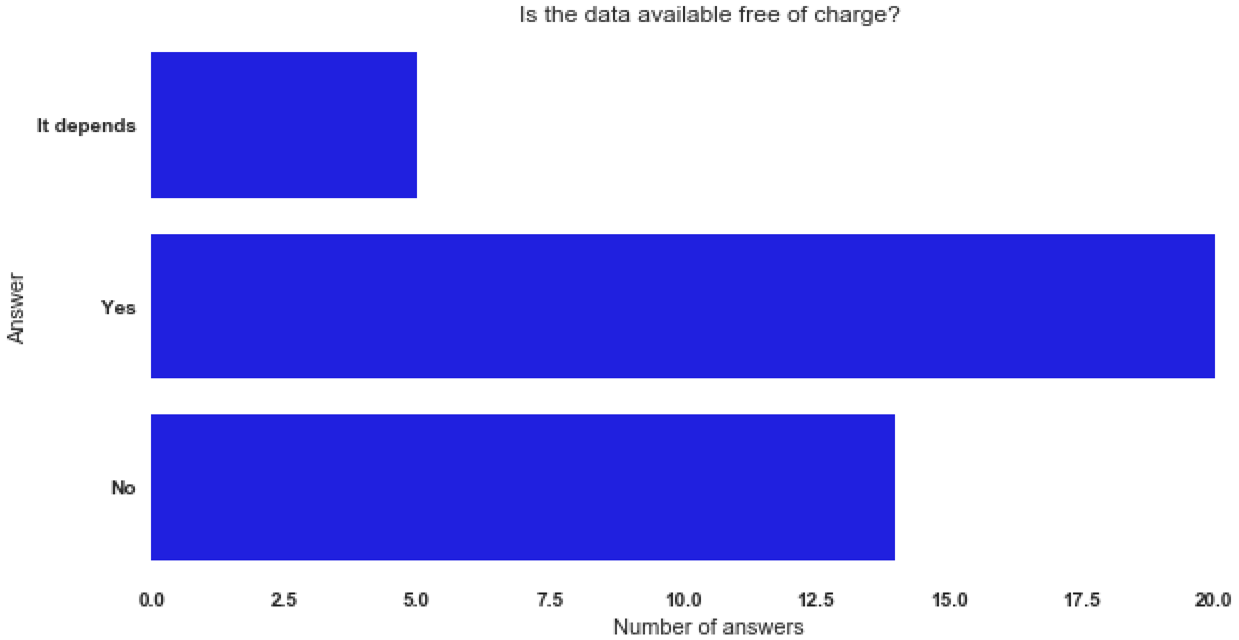

Figure 14: Is the data available free of charge?

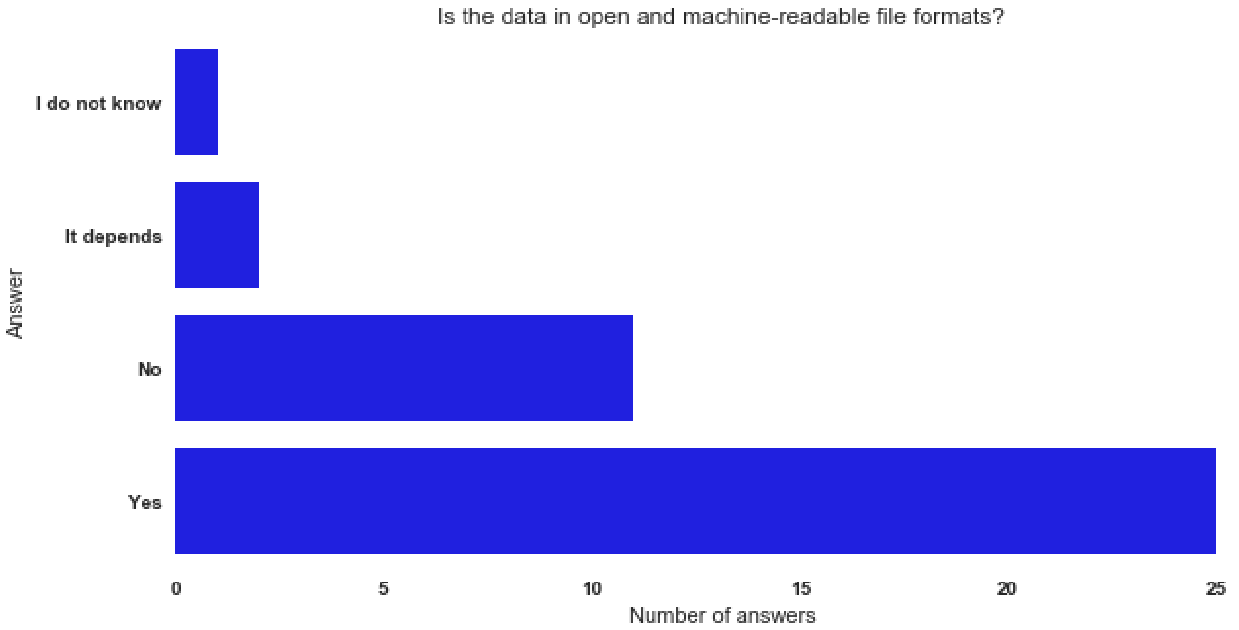

Figure 15: Is the data in open and machine-readable file format?

Figure 16: Is the data online without the need to register or request access to the data?

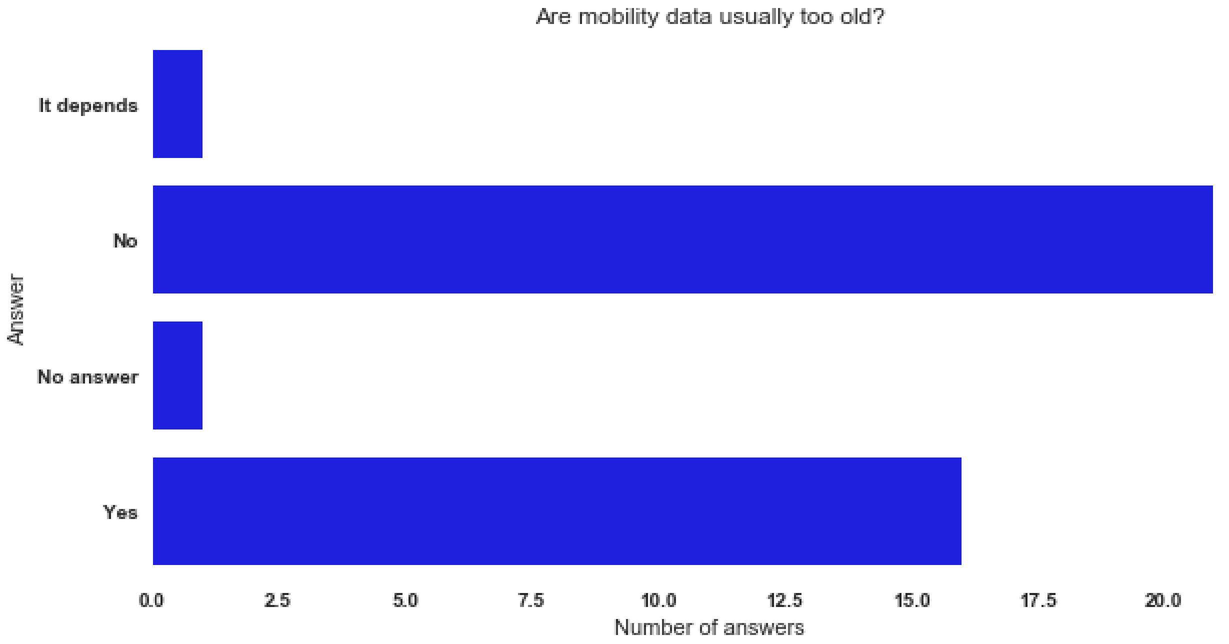

Figure 17: Are mobility data usually too old?

5. Business Models Acceptance

This section describes a dataset of the 36 data-sharing companies selected based on the market analysis reported in D3.2 Data Sharing Market Technological Developments. This dataset allows for studying the characteristics of the different value chains. We propose a methodology based on the Canvas model, a tool for capturing information on how companies organize themselves and create added value.

The Canvas model proposes nine categories representing the constituent elements of the company, namely the Customer Segment, Value Proposition, Channels, Customer Relationship, Revenue Streams, Key Resources, Key Partners and Key Activities and, finally, the Cost Structure.

Based on this classification, we associate a dummy variable and describe the characteristics of each company according to the Canvas model.

For instance, in the Canvas category called **Customer Segment**, we identified the customers the company might be interested in, such as individuals (average users or professional developers), companies and public institutions.

The **Value Proposition** represents the products and services offered by the company. In the data-sharing market, companies sell integrability with other platforms, free data, consulting services, marketplace or spatial analysis.

The **Channels** are how companies distribute their products or services, for instance, if the company organizes events for sales purposes or operates through social networks and other forms of communication like email, mobile apps or other direct sales channels.

The **Customer Relationship** are the elements through which the company establishes relationships with its customers and suppliers and integrates the data from the "Channels". More generally, it indicates how the company communicates with the outside world and establishes relationships (websites, user communities, personal accounts, training courses, support and monitoring, and product convenience).

The **Revenue Streams** generated by the corporate activities are divided based on how users use the services (subscriptions, licenses, commissions, personalized payments, etc.). The Key Resources identify the factors that allow activities to be carried out (databases, open-source resources, visualization tools, networks and know-how).

The **Key Activities** are the elements associated with the Value Proposition and indicate how to use the latter in the supply of goods/services (mapping, localization, API development, etc.).

The **Key Partners** are the subjects with whom the company identifies strategic collaborations. The Key Partners could be software suppliers, platforms and data, manufacturing industries and institutions (research institutions or public authorities such as government bodies, cities and local authorities).

Funded by the European Union

Finally, the **Cost Structure** identifies how the company supports its expenses (voluntary donations or pricing on platforms, datasets, services or data resale).

Alongside the dummies to describe the data-sharing companies according to the Canvas framework, we considered the qualitative variables identified in D3.2 to define the value chain generated by these companies. The first is the data source (i.e. whether the company shares data open, closed or limited-access). The second variable defines the relationships between the organization and its partners. The third variable defines the costs associated with the service management that can be variable, voluntary, free or imposed upon access. The fourth variable identifies the subjects who provide the data. Finally, the fifth variable enriches the information regarding the possible ways in which users sell or exchange data on the platforms.

Table 6 shows an extract of the dataset created based on Canvas. The first column reports the company, the second column shows the considered data-sharing product, and CS is the acronym for Customer Segment. For each customer segment, a dummy variable indicates the type of customers the company addresses. In this example, according to the information reported on the web, AKKA datahub is a product addressed to developers and professional users and manufacturers. Furthermore, the table includes the information reported in D3.4: data sources, maintenance responsibility, service maintenance costs, data providing/data analytic service responsibility and data reselling between users in the platform.

Table 6: *Example of dataset for the acceptance of the Business Models*

Name of the Company	Product	CS) AVG consumers	CS) Developers & professional users	CS) Mobility & transport operators	CS) Manufacturers (automotive, etc.)
AKKA technologies	AKKA datahub	0	1	0	1
Esri	ArcGis https://www.esri.com/en-us/arcgis/about-arcgis/overview	1	1	1	1
Carto	CARTO https://carto.com/	0	1	1	1
Enroute SAS	Chouette SaaS https://enroute.mobi/	1	1	1	0
Ckan	AKKA datahub	1	1	1	1

The sum of the dummy variables for each Canvas category allows the implementation of the agglomerative cluster analysis. Table 7 reports an example of an aggregated dataset.

Table 1

Funded by the European Union

Table 7: *Example of aggregated dataset for the acceptance of the Business Models*

Name of the Company	Product	Customer Segment (CS)	Channels (CH)	Customer Relationship (CR)	Value Proposition (VP)
AKKA technologies	AKKA datahub	5	5	5	6
Esri	ArcGis https://www.esri.com/en-us/arcgis/about-arcgis/overview	7	7	7	7
Carto	CARTO https://carto.com/	6	7	7	7
Enroute SAS	Chouette SaaS https://enroute.mobi/	7	7	7	7
Ckan	AKKA datahub	7	7	6	6

Table 8 shows the number of companies for each Business Model identified in D3.4. The data-sharing companies can be characterised based on a mix of BMs. Table 9 reports the new classification of companies obtained with the agglomerative cluster analysis implemented using the dataset reported in Table 7.

Table 8: *Number of companies for each Business Model*

BM code	# companies
BM1	4
BM3	1
BM4	4
BM6	2
BM7	1
Mixed	24
Total	36

Funded by the European Union

Table 9: Number of companies for each cluster

Cluster	# companies
cluster 1	11
cluster 2	6
cluster 3	9
cluster 4	2
cluster 5	7
cluster 6	1
Total	36

The Canvas components analysis of each cluster allows the classification of the data-sharing companies. The objective is to overcome the limitation of categorisation reported in Table 8, where 68% of the companies are in the category "Mixed".

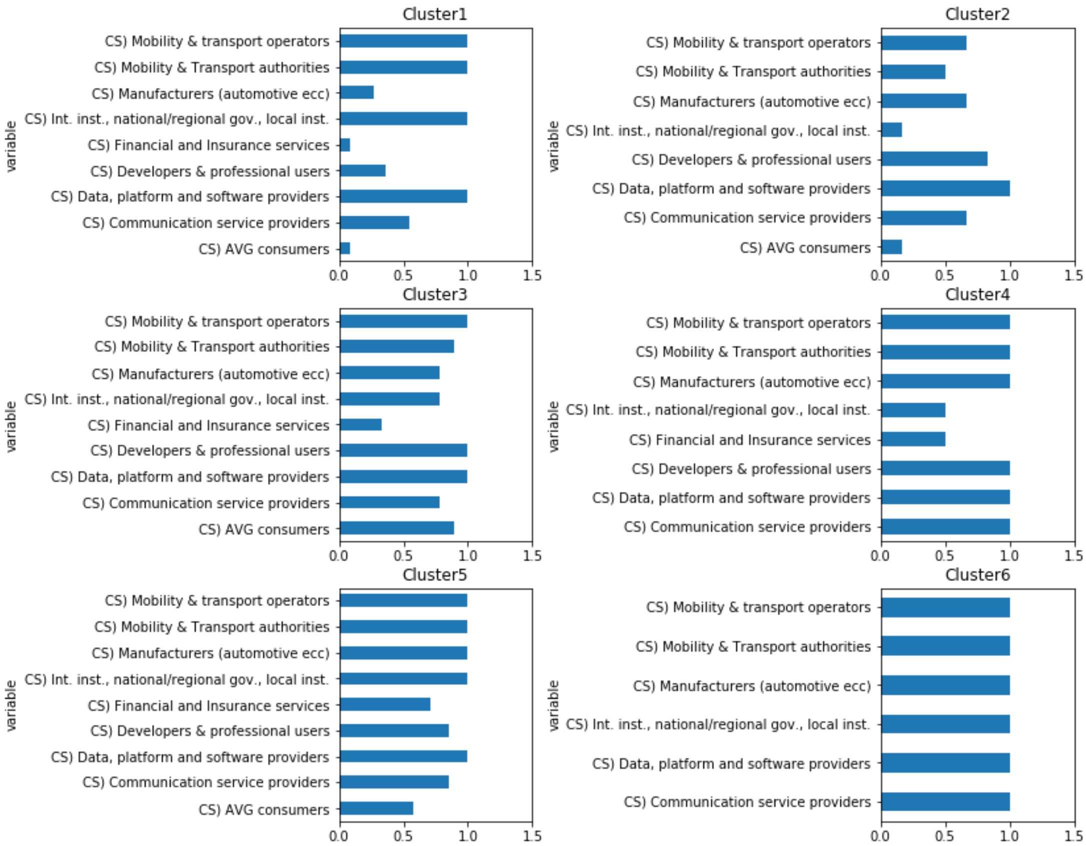

Figure 18: Customer segment (CS) of each cluster of data sharing companies

Figure 18 shows the dummy variables of the clusters of data-sharing companies. A different composition of customers characterises each class. The higher the number of customers in each segment, the more companies deal with these customers. Cluster 1 and Cluster 2 are similar in terms of customer segment although more companies of Cluster 2 address their services to manufacturers and companies of Cluster 1 to International institutions, national or regional governments and local institutions. All the considered data-sharing companies offer their services to data, platform and software providers and mobility and transport operators.

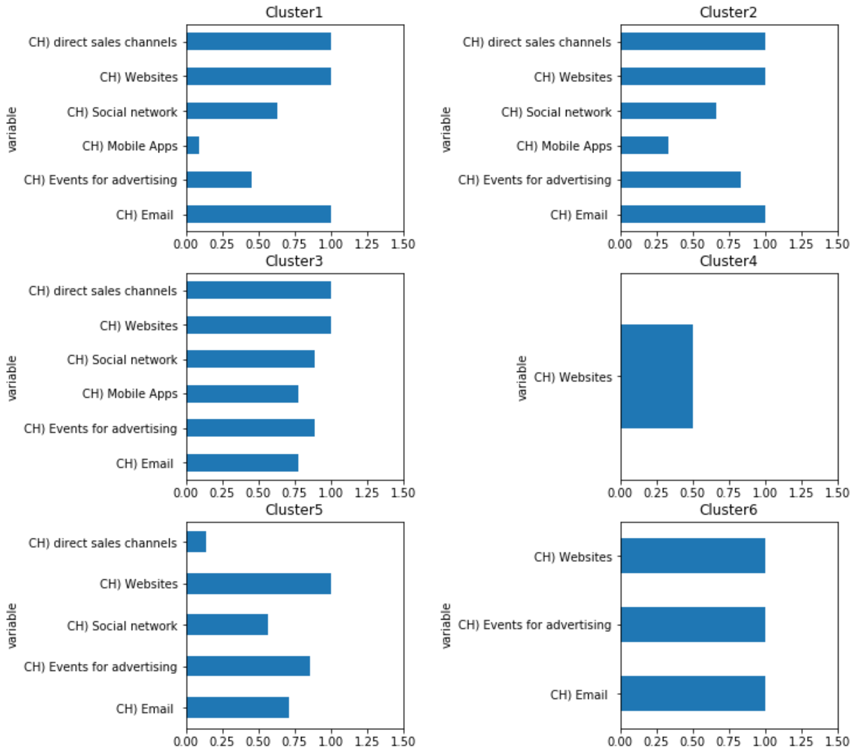

Figure 19: *Channels of each cluster of data sharing companies*

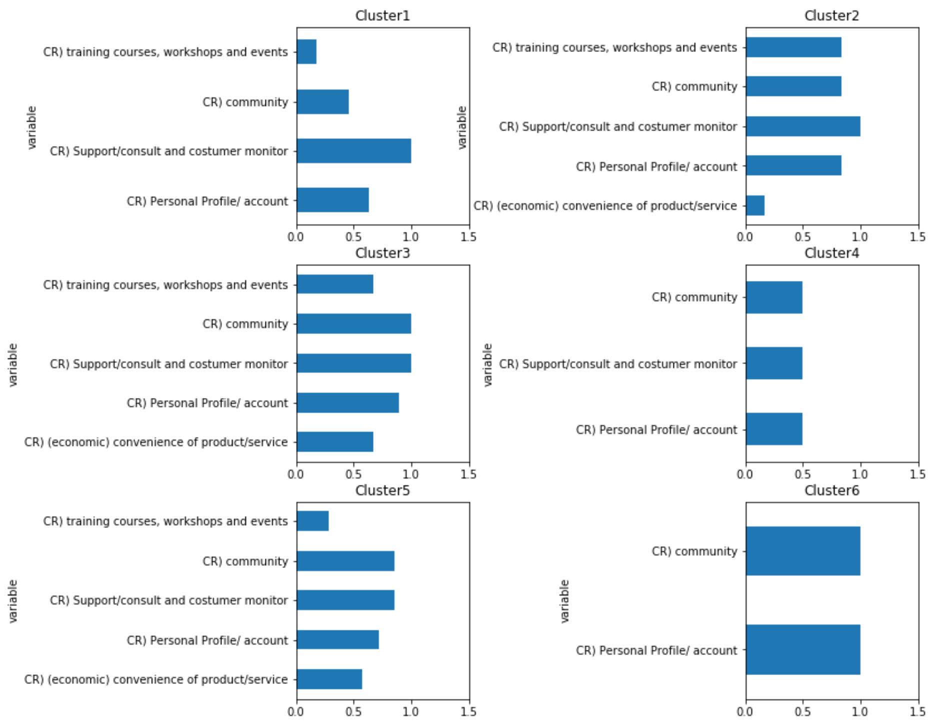

Figure 20: Customers Relationships (CR) of each cluster of data sharing companies

Figure 2 presents the Channels (CH) of each cluster. Cluster 4 has only one company which has only one channel (website). Cluster 1, Cluster 2 and Cluster 3 have a direct sale channel and sell through the website. All companies of Cluster 1, Cluster 2 and Cluster 6 also have an email channel. The companies of Cluster 1 and Cluster 3 distribute their product using the social network.

Figure 20 presents the customer relationship (CR) of each cluster. Cluster 6 has only one and provides services to the community through a personal profile or user account. Clusters 1, 2 and 3 are more heterogeneous in terms of the customer relationship.

Figure 21 shows the key activities (KA) of each cluster. Clusters 1, 2, 3 and 6 offer, among other things, data management and storage services, open data-sharing services, and services for the transport sector, such as ticketing and mapping.

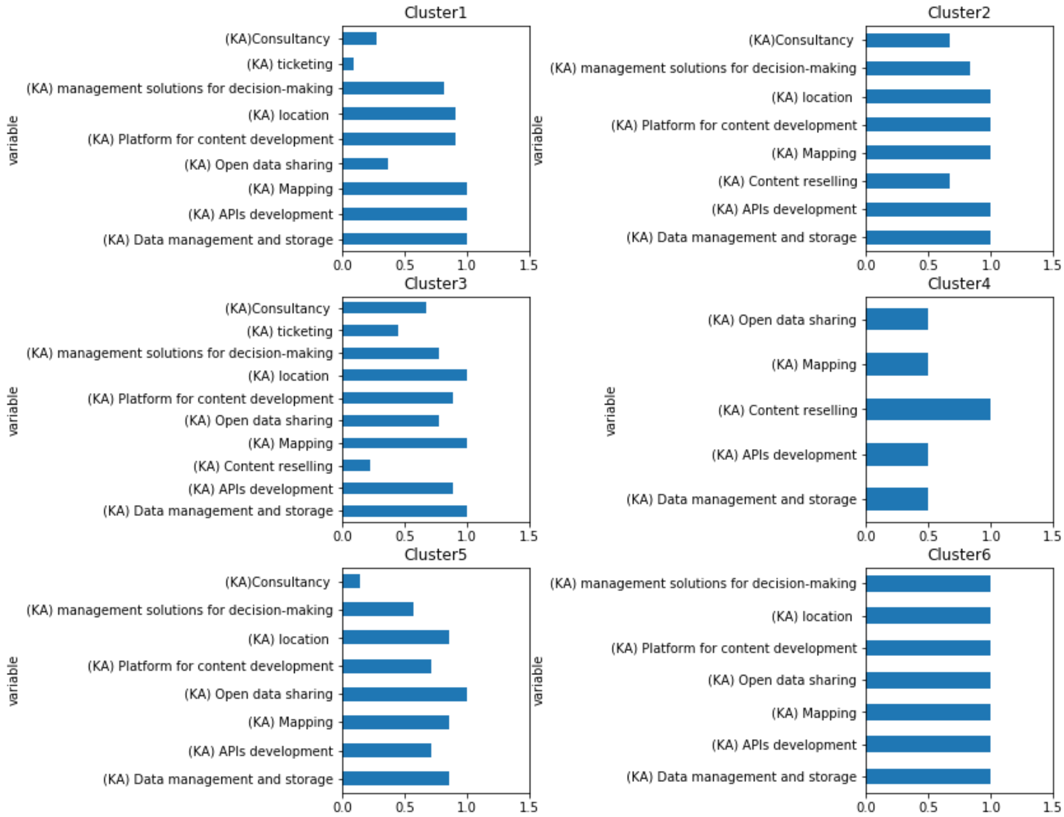

Figure 21: Key Activities (KA) of each cluster of data sharing companies

Figure 22 evidences that the companies of Cluster 1 mainly work with public authorities. Cluster 1, Cluster 2 and Cluster 3 work with several different partners, although the companies of Cluster 2 work less with No profit organizations and industries such as OEMs. The key partners of Cluster 2 are communication, software, platform and data providers. The key partners of Cluster 4 are public authorities, logistics operators, businesses and data providers.

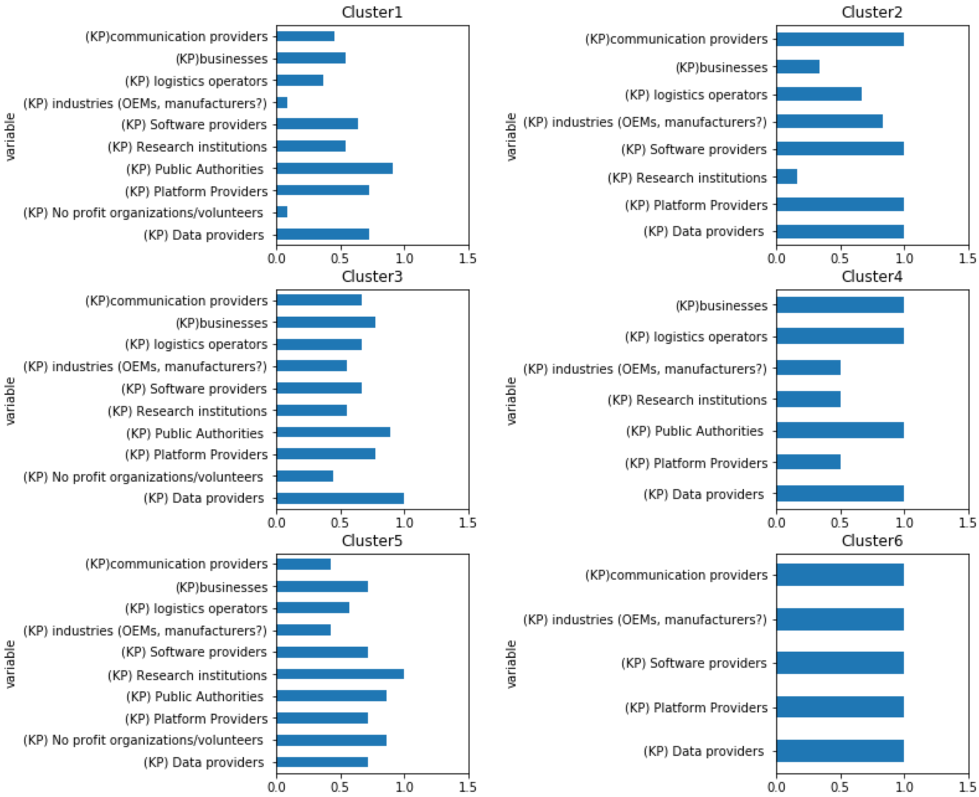

Figure 22: Key Partners (KP) of each cluster of data sharing companies

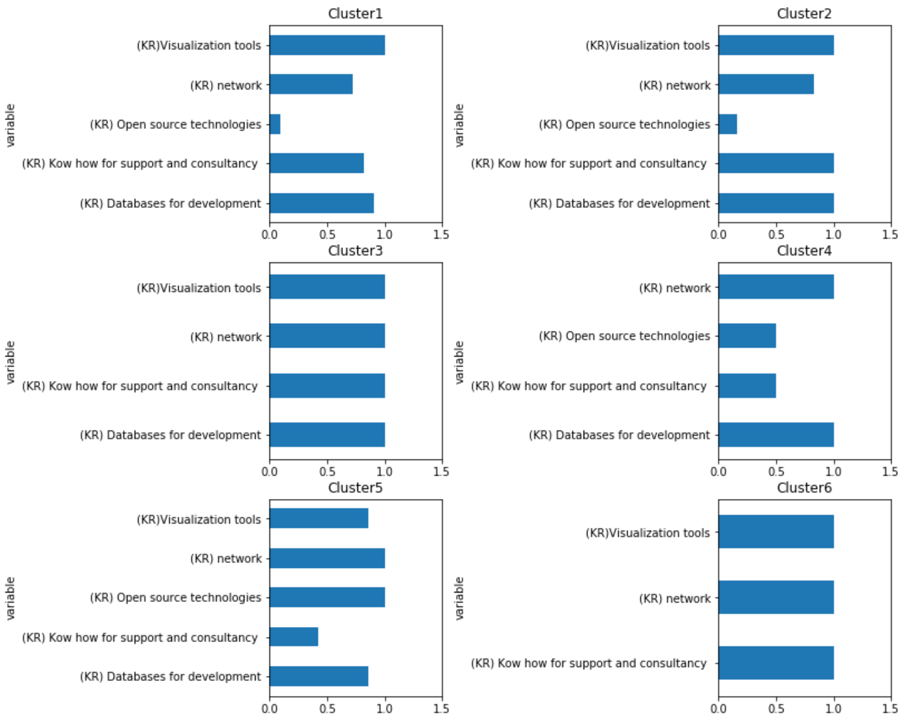

Figure 23: Key Resources (KR) of each cluster of data sharing companies

Figure 23 shows the key resources (KR) of each cluster. All clusters except the number 4 provide visualization tools. Furthermore, all clusters except the Cluster 6 provide databases for development. Know-how and consultancy services are the key resources of Cluster 1, 2, 3 and 6.

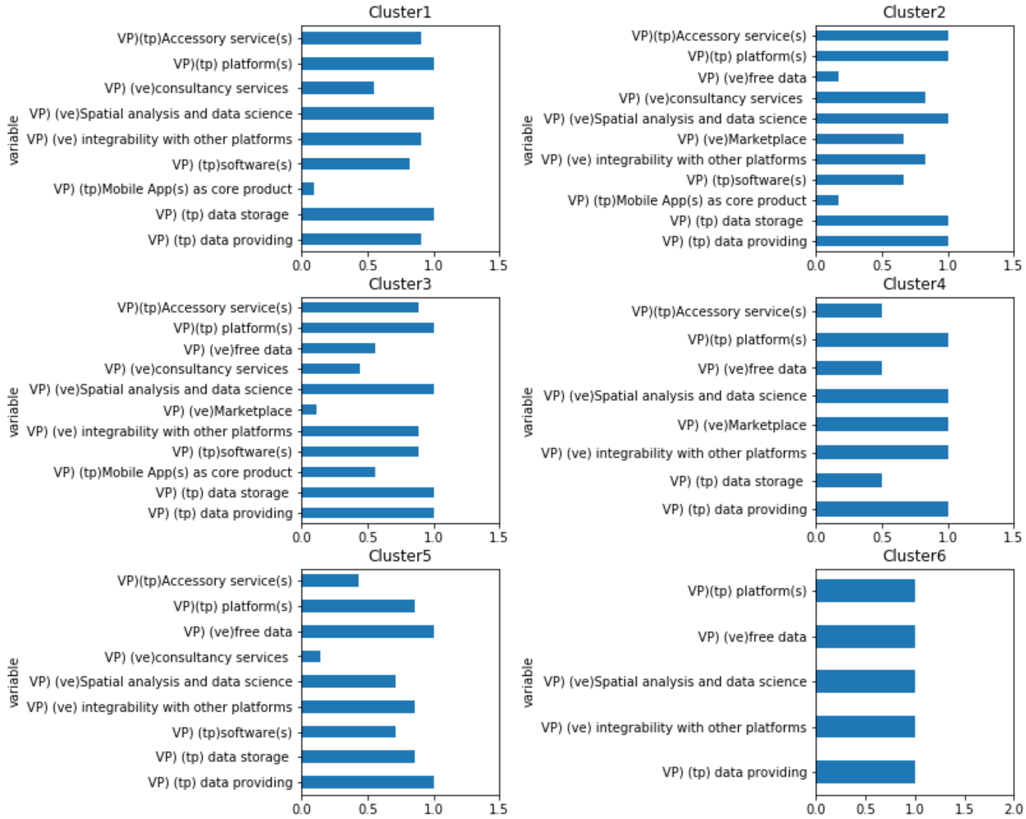

Figure 24: Value Proposition (VP) of each cluster of data sharing companies

The value proposition (VP) of Figure 24 shows that all companies provide platform because that was the criteria for their selection. Free data are offered by the companies of Cluster 4. All companies expect some companies of Cluster 5 offer spatial analysis and data science services.

Funded by the European Union

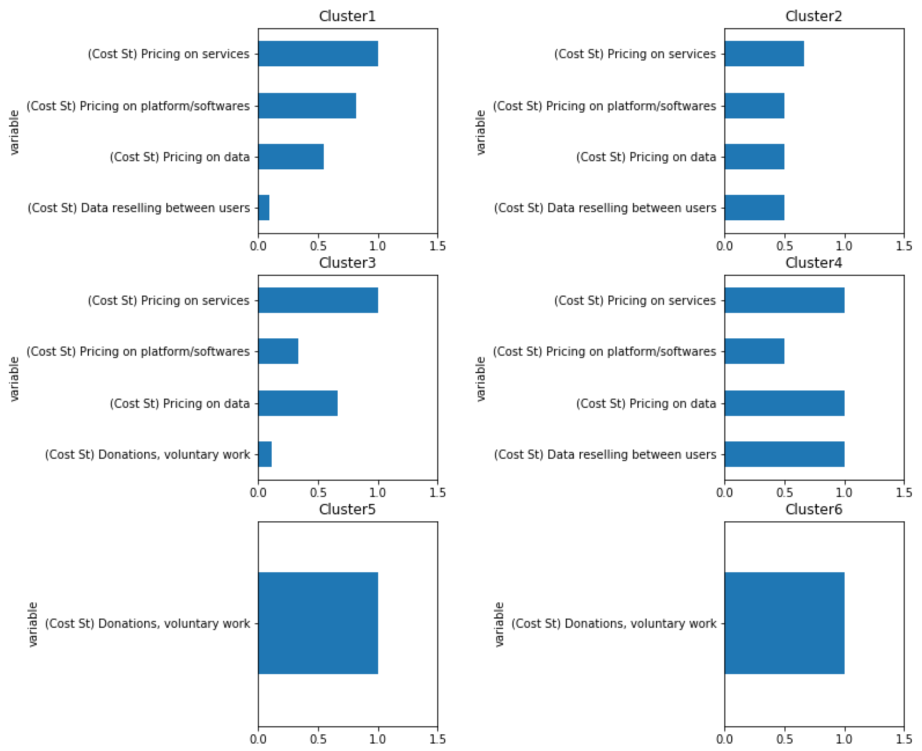

Figure 25: Cost Structure (Cost St) of each cluster of data sharing companies

Surprisingly, Figure 25 shows the cost structure of Cluster 1, formed by donations, voluntary work and non-profit partnerships. Cluster 2 and Cluster 3 have very similar Canvas components. Furthermore, the main difference between Cluster 2 and Cluster 3 is that the latter monetises data and services.

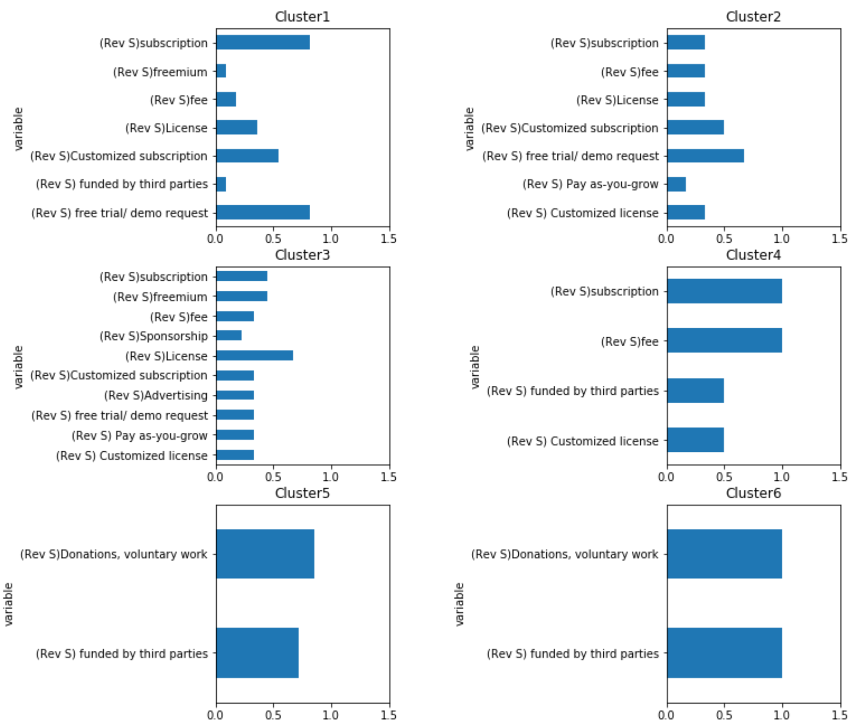

Figure 26: Revenue Stream (Rev S) of each cluster of data sharing companies

Funded by the
European Union

Figure 26 presents the revenue stream of each cluster. Cluster 1 and Cluster 3 are similar in terms of revenues derived from donations, voluntary work and third parties. The revenue streams of Cluster 3 are more complex because they derive from many sources.

6. Conclusion, next steps and future work

Despite we are living in the Age of Data (MacAfee et al., 2012), there are still many challenges to be solved before we get to talk about a shared culture of data and to reach awareness of the importance of data and the value associated with it. There is still little knowledge of the necessary organizational aspects

The objective of this analysis was to have a deeper understanding of the data-sharing culture. More specifically, this work has explored the relationship between data sharing, trust between partners and social impact, filling some gaps in the literature that emphasize how awareness of the importance of data does not translate into an understanding of how to pass from data to a container of value. Understanding the forms of collaboration that can encourage the exchange of data in a partnership, as well as the understanding of actions that can inspire greater openness towards the exchange of data, would also lead to an academic and industrial reality ready to deeply live the Age of data (MacAfee et al., 2012).

The analysis presented in this deliverable shows how the textual data of the interviews of different actors in the transport sector with argumentation mining and topic modelling can define principles of data-sharing culture. In particular, the research used argumentation mining and topic modelling to extract the interviewee's claims and then the qualitative analysis allowed to explain several topics. Among the others, we have identified a view on prerequisites that facilitate data sharing, on the relevant social impact of data sharing, on the concept of the ecosystem of data sharing and, finally, on a necessary vision that can guide future activities.

As for the next step, we will carry out 29 additional interviews with different actors of the data sharing culture. The idea is to compare the topic modelling with the grounded theory to analyse the collected interviews and analyse the collected interviews to enrich the findings, allowing a deep analysis (Baumer et al., 2017). Therefore, it would be interesting to deploy a similar qualitative approach based on the grounded theory (Strauss and Corbin, 1997) and to classify the written material into the identified categories of similar meanings (Moretti at al., 2011), making comparisons between empirical data and concepts and between concepts and categories (Gregory, 2011).

Once the analysis with the grounded theory will be performed, it will be possible not only to enrich the results presented in this document but it will be also possible to test and verify the real efficiency of argumentation mining and topic modelling analysis.

Furthermore, this deliverable aimed to classify data-sharing companies' Business Models based on the Canvas classification. In the context of T5.2 activities, we plan to develop a composite index starting from the subdivision of the qualitative variables associated with each Canvas category. The index will measure the greater or lesser degree of openness towards data sharing. The underlying idea is that this openness is not exclusively associated with the company's choice to provide free data but also with the company's relations with customers, communication and distribution channels,

Funded by the European Union

target users, partners, etc. The BM index will associate a specific weight to each qualitative variable relating to the Canvas categories. Therefore, it will be possible to rank the resulting values on a scale. The correlation analysis between companies' characteristics will allow to associate weights according to the opening towards data sharing. The underlying hypothesis that companies that are totally open to data sharing adopt choices that involve all categories of users to access data regardless of the most represented class of users, who communicate and distribute the products through as many channels as possible and give the opportunity to collaborate in the supply of goods and services to as many stakeholders as possible.

Finally, the survey on the Transport Cloud acceptance showed that the users are willing to use the Transport Cloud platform. This intention will match the actual behaviour if the Transport Cloud is easy to use, has a fast response time and allows to find the data the user needs. As for the next step, we plan to extend the survey and propose innovative methodologies to analyse the results.

7. References

Al-Zahrani, Fahad Ahmad. “Subscription-Based Data-Sharing Model Using Blockchain and Data as a Service.” IEEE Access 8: 115966-115981 (2020)

Barde, B. V. and Bainwad, A. M. "An overview of topic modeling methods and tools," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 745-750 (2017)

Bauer, P.C., Keusch, F., Kreuter, F: Trust and cooperative behavior: Evidence from the realm of data-sharing. PloS one 14(8), e0220115 (2019)

Baumer, E.P., Mimno, D., Guha, S., Quan, E., Gay, G.K: Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? Journal of the Association for Information Science and Technology 68(6), 1397–1410 (2017)

Blei, D.M: Probabilistic topic models. Communications of the ACM 55(4), 77–84 (2012)

Brea Solís, H., Casadesus Masanell, R., & Grifell Tatjé, E. (2015). Business Model Evaluation: Quantifying Walmart's Sources of Advantage. Strategic Entrepreneurship Journal, 9(1), 12-33.

Carmichael, P.: Not just about gadgets: Habit, innovation and change in the design of learning technologies. E-Learning and Digital Media 12(3-4), 279–294 (2015)

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. Advances in neural information processing systems (pp. 288–296). . Retrieved from http://machinelearning.wustl.edu/mlpapers/paper_files/NIPS2009_0125.pdf.

Chen, W., Chen, Y., Chen, X., Zheng, Z: Toward secure data sharing for the iov: A quality-driven incentive mechanism with on-chain and off-chain guarantees. IEEE Internet of Things Journal 7(3), 1625–1640 (2019)

Covington, M.A: A fundamental algorithm for dependency parsing. In: Proceedings of the 39th annual ACM southeast conference. vol. 1. Citeseer (2001)

Davis, F. D. : Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, 319-340, (1989)

Du, T.C., Lai, V.S., Cheung, W., Cui, X: Willingness to share information in a supply chain: A partnership-data-process perspective. Information & Management 49(2), 89–98 (2012)

M., Scholl, H.J., Wimmer, M.A., Bannister, F. (eds) Electronic Government. EGOV 2014. Lecture Notes in Computer Science, vol 8653. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44426-9_21

El Sawy, O. A., & Pereira, F. Business modelling in the dynamic digital space: An ecosystem approach. Heidelberg: Springer (2013)

Figueiredo, A.S: Data sharing: convert challenges into opportunities. Frontiers in public health 5, 327 (2017)

Gellerman, H., Svanberg, E., Barnard, Y: Data sharing of transport research data. Transportation Research Procedia 14, 2227–2236 (2016)

Giessmann, A., & Legner, C. Designing business models for cloud platforms. Information Systems Journal, 26(5), 551-579, (2016).

Gordijn, J., & Akkermans, H. (2001). Designing and evaluating e-business models. IEEE intelligent Systems, 16(04), 11-17.

Gregory, R.W: Design science research and the grounded theory method: Characteristics, differences, and complementary uses. In: Theory-guided modeling and empiricism in information systems research, pp. 111–127. Springer (2011)

Hagen, L, Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models?, Information Processing & Management, 546, 1292-1307 (2018)

Hagen, L., Uzuner, O., Kotfila, C., Harrison, T. M., & Lamanna, D. Understanding citizens' direct policy suggestions to the federal government: A natural language processing and topic modeling approach. 2015 48th Hawaii international conference on system sciences (HICSS) 2134–2143, (2015) https://doi.org/10. 1109/HICSS.2015.257

Honnibal, M., Montani, I: spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing (2017)

Hopkins, D. J., & King, G. A method of automated nonparametric content analysis for social science. American Journal of Political Science, 54(1), 229–247 (2010).

Keseru, I., Coosemans, T., Macharis, C: Stakeholders' preferences for the future of transport in europe: Participatory evaluation of scenarios combining scenario planning and the multi-actor multi-criteria analysis. Futures 127, 102690 (2021)

Kulviwat, S., Bruner II, G. C., Kumar, A., Nasco, S. A., & Clark, T. (2007). Toward a unified theory of consumer acceptance technology. Psychology & Marketing, 24(12), 1059-1084.

Lau, B.P.L., MarAKKAlage, S.H., Zhou, Y., Hassan, N.U., Yuen, C., Zhang, M., Tan, U.X.: A survey of data fusion in smart city applications. Information Fusion 52, 357–374 (2019)

Li, S., Lin, B: Accessing information sharing and information quality in supply chain management. Decision support systems 42(3), 1641–1656 (2006)

Lind, M., Watson, R.T., Ward, R., Bergmann, M., Bjorn-Andersen, N., Rosemann, M., Haraldson, S., Andersen, T: Digital data sharing: The ignored opportunity for making global maritime transport chains more efficient. Unctad Transport and Trade Facilitation Newsletter (2018)

Lippi, M., Antici, F., Brambilla, G., Cisbani, E., Galassi, A., Giansanti, D., Magurano, F., Rosi, A., Ruggeri, F., Torroni, P.: Amica: an argumentative search engine for covid-19 literature. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI. vol. 22 (2022)

Lippi, M., Torroni, P: Margot: A web server for argumentation mining. Expert Systems with Applications 65, 292–303 (2016)

Liu, Y., & Xu, S. A local context-aware LDA model for topic modeling in a document network. Journal of the Association for Information Science and Technology, 68(6), 1429–1448, (2017)

McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D., Barton, D.: Big data: the management revolution. Harvard business review 90(10), 60–68 (2012)

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Neural and Information Processing System (NIPS) (2013), https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf

Moretti, F., van Vliet, L., Bensing, J., Deledda, G., Mazzi, M., Rimondini, M., Zimmermann, C., Fletcher, I.: A standardized approach to qualitative content analysis of focus group discussions from different countries. Patient education and counseling 82(3), 420–428 (2011)

Moschitti, A.: Efficient convolution kernels for dependency and constituent syntactic trees. In: European Conference on Machine Learning. pp. 318–329. Springer (2006)

Moschovou, T., Vlahogianni, E., Rentziou, A.: Challenges for data sharing in freight transport. Advances in Transportation Studies 48, 141–152 (2019)

Munoz-Arcentales, A., López-Pernas, S., Pozo, A., Alonso, Á., Salvachúa, J., Huecas, G.: An architecture for providing data usage and access control in data sharing ecosystems. Procedia Computer Science 160, 590–597 (2019)

Osterwalder, A., & Pigneur, Y. Business model generation: a handbook for visionaries, game changers, and challengers (Vol. 1). John Wiley & Sons (2010)

Paulsson, A., Isaksson, K., Sørensen, C.H., Hrelja, R., Rye, T., Scholten, C.: Collaboration in public transport planning–why, how and what? Research in transportation economics 69, 377–385 (2018)

Porter, M.E. Competitive Advantage – Creating a Sustaining Superior Performance, The Free Press, New York (1985)

Rappa, M. A. The utility business model and the future of computing services. IBM systems journal, 43(1), 32-42 (2004)

Rehurek, R., Sojka, P.: Gensim–python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic 3(2) (2011)

Rouhani, S., Deters, R.: Data trust framework using blockchain technology and adaptive transaction validation. IEEE Access 9, 90379–90391 (2021)

Rouibah, K., Ould-Ali, S.: Dynamic data sharing and security in a collaborative product definition management system. Robotics and Computer-Integrated Manufacturing 23(2), 217–233 (2007)

Seshadri, K., Mercy Shalinie, S., & Kollengode, C. Design and evaluation of a parallel algorithm for inferring topic hierarchies. Information Processing & Management, 51(5), 662–676, (2015) https://doi.org/10.1016/j.ipm.2015.06.006.

Shrestha, A. K., & Vassileva, J. (2019, December). User acceptance of usable blockchain-based research data sharing system: an extended TAM-based study. In 2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA) (pp. 203-208). IEEE. F. D. Davis, A Technology Acceptance Model for Empirically Testing New End-User Information Systems, 1986.

Simanis, E., Hart, S.: Innovation from the inside out. Top 10 Lessons on the New Business of Innovation 9 (2011)

Strauss, A., Corbin, J.M.: Grounded theory in practice. Sage (1997)

Suominen, A., & Toivanen, H. Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. Journal of the Association for Information Science and Technology, 67(10), 2464–2476 (2016)

Teece, D. J.. Business models, business strategy and innovation. Long range planning, 43(2-3), 172-194 (2010).

Venkatesh, V., L. Thong, J. Y., & Xu, X. Consumer acceptance and use of information technolgy: Extending the unified theory of acceptance and use of technology. MIS Quarterly, 36(1), 157-178 (2012)

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425-478 (2003)

Williams, M. D., Rana, N. P., & Dwivedi, Y. K. The unified theory of acceptance and use of technology (UTAUT): a literature review. Journal of enterprise information management . (2015)

Witkowski, K.: Internet of things, big data, industry 4.0–innovative solutions in logistics and supply chains management. Procedia engineering 182, 763–769 (2017)

Zhao, X., Jin, P., & Yue, L.. Discovering topic time from web news. Information Processing & Management, 51(6), 869–890 (2015) https://doi.org/10.1016/j.ipm.2015.04.001.

Funded by the European Union

8. Annexes

8.1. Annex I

Updated INTERVIEW PROTOCOL:

COLLABORATION EFFECTIVENESS FOR QUALITY DATA SHARING, SOCIAL AND ENVIRONMENTAL IMPACT IN THE MOBIDATALAB FRAMEWORK

INTRODUCTION:

Dear, first of all, thanks for your time: I know we are overwhelmed with confcall and your availability for this interview is very precious.

So, as anticipated via mail, the focus of the interview is the data sharing in the mobility transport sector. The interview is carried out in the context of MobiDatalab, a European union funded project that aims to create a data-sharing culture to let transport authorities, operators and other mobility stakeholders in Europe know about the advantages of sharing data. The final goal of MobiDataLab is to develop knowledge and a cloud solution to easing the sharing of data.

This interview will last around 1 hour: we need your opinions and your experience in order to understand the real needs regarding data and the data sharing culture. There will be 4 sections:

- SECTION A aims to assess the collaboration and quality of knowledge sharing
- SECTION B aims assess data sharing and collaboration culture
- SECTION C focuses on the social impact of the data-sharing culture

If it is ok with you, I will record this interview: I needed it only for research purposes in order to enable transcription of data and ensure that no info is lost in the process of data gathering and analysis. I want to underline that the data will be analysed in an aggregate form, and that single responses and sources will remain anonymous.

Just a short introduction from my side. My name .., I work with Selini in the MobidataLab project. We are part of ICOOR an interuniversity consortium that focuses mainly on Europen Projects within different topics as data sharing, 5G, autonomous vehicles, electric vehicles, logistics and urban mobility. Within MobidataLab we are mainly involved in the evaluation of the solutions proposed in the project and in the collection of opinions from the stakeholders.

So, do you have any questions?

Ok so, if you are ready, we can start with the first section.

<u>SECTION A: COLLABORATION AND QUALITY OF KNOWLEDGE SHARING</u>:

First of all, I would like to ask you a couple of questions related to collaborations for data mobility integration.

- Can you tell me a bit about your organization's approach to data sharing?
- What do you think are the main opportunities and risks of sharing data for integrated mobility?
- How do you balance risks and opportunities, what do you think is the key?

- Who are the stakeholders (people and organizations) that you collaborate the most with in terms of sharing/receiving mobility data?
- Do you think that stakeholders of mobility data exchange share common goals? Or do you think there are different goals?
- (if there are different goals) Do you see them as complementary?
- Would you say that you are satisfied with how they collaborate to provide the data? Why? What could be improved? What would you change? If you have super powers, what would you do to improve data sharing? What is missing? What are the critical points?
- Are they contributing to developing standards and compliance mechanisms in the present or will they do so in the future?
- Do you think that previous collaboration/knowledge sharing on mobility is a prerequisite for developing successful solutions for this project? Which are the most important prerequisites for you?
- In your opinion what is the difference between data sharing and data knowledge?
- Would you be more open to share knowledge or data? Why?
- Do you think a forerunner would be needed for best practice in data sharing? Why that one couldn't be your company?
- Do you think reassurance would be needed? Would you appreciate a specific list with how/from whom the data that you shared are used?

Regarding social impact and environmental goals:

- Do you think that different actors in the network such as dispatchers, transport customers, fleet owners, policymakers, integrated mobility service providers, mobility researchers, etc. have different social and environmental goals or do you think that they are pretty aligned?
- Overall, would you say that your organization and the other stakeholders have a shared vision of integrated mobility?
- Would you say the data sharing has opportunities on social impact? How?

Would you say that you **trust** the stakeholders that you collaborate with?

- do you find them honest, reliable, and benevolent in attaining the project's goal (i.e., data sharing for integrated mobility)?
- What would you like to change in your relationship?
- How can the trust be improved?
- Do you think incentives could be useful for data sharing? From whom should they come? For your specific situation which incentives would you like to have?

It is often acknowledged that **data quality** is one of the most important factors of journey planning. However, what people mean by data quality may vary.

- What is data quality to you and what are the main data quality challenges that you see for sharing data quality?
- Would you say that you are satisfied with the quality of the data they share?

If needed: mention completeness, consistency, accuracy and integrity of datasets only if answers are too generic or elusive.

- What do you think are the key factors for sharing high quality data?
- How much does collaboration among parties' matter, and how much can data analytics and machine learning techniques help increase data quality? Do you think the second can supplement or even substitute the first?
- How can data analytics and learning techniques help to reduce the impact of transport on the environment, in your view?

Funded by the European Union

- Do you think that data analytics and learning can be of help to create more inclusive digital transport services?
- What about the ability to interpret and use data for decision making?
- With advanced data analytics and ML, also the need for expert knowledge from researchers and data scientists' increases. Do you see that compatible with the mobility sector?
- Do you think it can have some positive aspects on society and employment?
- How do you see that relating to companies and organizations current best practices for data analysis? How could the two be integrated?
- Regarding the data type: would you say that certain data are easier to be shared? Why?
- Do you think that increasing the data quality it would be possible to increase the opportunities related to data?

Are you involved in a particular project regarding data sharing? Would you like to? What would you like to focus on?

SECTION B: QUESTIONS TO SPECIFICALLY ASSESS DATA SHARING AND COLLABORATION CULTURE

Let's talk about the Relationship you have with other realities that you work for

- How much do you currently work on continuously improving info sharing with your partners/?
- How much do you know about the data management practices of your partners?
- How much have you learned from the data management practices of your partners in the last years?
- Do you and your partners exchange knowledge and viewpoints on a regular basis about data mobility and possible methods of integration?
- Are there any formal or structured initiatives such as periodic training courses, workshops and seminars where you can share knowledge and experience on data mobility with your partners?
- Do you think that you and your partners have complementary resources?
- Do you consider yourself or your company a pioneer? Is it a role that you like? Or is it a role that you would like to have?
- Which are the main reason to don't collaborate with a partner?
- Is there a specific category of stakeholders that you trust the most?
- Do you use a specific platform or a specific ecosystem to share data?

SECTION C: SOCIAL IMPACT

Let's talk about the **social aspect** now:

- What is social change to you and what kind of social change can integrated mobility solutions bring to the table?

 Only if they are too broad and generic follow up with the following:

Impact on What: Jobs, Income, Skills, Knowledge, Education, Awareness, Access and equality, Engagement, Accountability, Empowerment, Integrity? other? -for any of these, HOW?

Impact on Who: Participants in the project? Target groups of use cases at large? Wider community and society? Policy making?

Funded by the European Union

- Do you think that recent exogenous shocks such as the COVID-19 pandemic and energy crisis due to the Russia-Ukraine conflict can foster a greater attention to MaaS and journey planning tools?

MobiDataLab consortium

The consortium of MobiDataLab consists of 10 partners with multidisciplinary and complementary competencies. This includes leading universities, networks and industry sector specialists.

@MobiDataLab
#MobiDataLab

https://www.linkedin.com/company/mobidatalab

For further information please visit **www.mobidatalab.eu**

MobiDataLab is co-funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879).