MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

D3.5 Societal and Environmental Impacts of Data Sharing Assessment Framework

23/01/2023

Author(s): Selini Hadjidimitriou (ICOOR)

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

Summary sheet

Deliverable Number	D3.5
Deliverable Name	Societal and Environmental Impacts of Data Sharing assessment framework
Full Project Title	MobiDataLab, Labs for prototyping future Mobility Data sharing cloud solutions
Responsible Author(s)	Selini Hadjidimitriou (ICOOR)
Contributing Partner(s)	AKKA, POLIS, AETHON, HERE
Peer Review	HERE, KUL
Contractual Delivery Date	30-04-2022
Actual Delivery Date	29-04-2022
Status	Final
Dissemination level	Public
Version	V1.0
No. of Pages	54
WP/Task related to the deliverable	WP3/T3.5
WP/Task responsible	ICOOR
Document ID	MobiDataLab-D3.5-SocietalEnvironmentalImpactsDataSharing AssessmentFramework-v1.0
Abstract	This document describes the evaluation methodology to assess the impact of data sharing services provided by the new Transport Cloud and the related Business Models on society and the environment.

Legal Disclaimer

Project partners

Organisation	Country	Abbreviation
AKKA I&S	France	AKKA
CONSORZIO INTERUNIVERSITARIO PER L'OTTIMIZZAZIONE E LA RICERCA OPERATIVA	Italy	ICOOR
AETHON SYMVOULI MICHANIKI MONOPROSOPI IKE	Greece	AETHON
HOVE	France	HOVE
HERE GLOBAL B.V.	Netherlands	HERE
KATHOLIEKE UNIVERSITEIT LEUVEN	Belgium	KUL
POLIS - PROMOTION OF OPERATIONAL LINKS WITH INTEGRATED SERVICES	Belgium	POLIS

Funded by the European Union

Document history

Version	Date	Organisation	The main area of changes	Comments
0.1	15/12/2021	ICOOR	ToC	
0.2	16/3/2022	ICOOR	The first draft of the deliverable	
0.3	4/4/2022	ICOOR	Integrating comments received from partners	
0.4	11/04/2022	ICOOR	Final version	
0.5	12/04/2022	ICOOR	Correct template	
0.6	20/4/2022	ICOOR	Integration of additional questions for the interviews and integration of KUL comments and review	
0.7	21/04/2022	ICOOR	Grammar checking	
0.8	26/04/2022	ICOOR	Adjustments after HERE peer review	
0.9	27/04/2022	ICOOR	Document ready for final quality check	
1.0	29/04/2022	AKKA	Ready for submission	

Executive Summary

This document describes the evaluation methodology to assess the impact of the data sharing culture on society and the environment. More specifically, this document presents the approach to evaluate the business models and the MobiDataLab Transport Cloud acceptance.

The first part of the deliverable describes, theoretically, the areas of impact of the data sharing culture. Furthermore, the first part includes a selection of digital transport services and how they can impact society or the environment. Concerning the MobiDataLab use cases described in D2.9, there are some hypotheses on their impacts on society and the environment. Based on these hypotheses, it was possible to formulate several open questions to assess the use cases and the data sharing culture's impacts on the economy, environment, and society. The existing projects and initiatives have allowed us to identify a set of general KPIs. These KPIs aim to measure the acceptance of the Transport Cloud by different types of users.

The second part of the deliverable deals with the business models acceptance. The evaluation methodology is the same as the one already adopted in D3.4 that describes the business models. Finally, the Unified Theory of acceptance and use of technology (UTAUT) and the non-functional requirements of the Virtual Labs, which are prototypes services, have allowed to set up the survey to evaluate the Transport Cloud acceptance. As in the case of the business models survey, the interviews will be carried out during MobiDataLab events and integrated with additional interviews to ensure the representativeness of the respondents' sample.

Table of contents

List of figures

List of tables

Funded by the European Union

Abbreviations and acronyms

Abbreviation	Meaning
AI	Artificial Intelligence
BM	Business Models
ETA	Estimated Time of Arrival
FAIR	Findable, Accessible, Interoperable and Reusable
GDP	Gross Domestic Product
GTFS	General Transit Feed Specification
IoT	Internet of Things
KPI	Key Performance Indicators
MaaS	Mobility as a Service (MaaS)
OBU	On Board Units
RSU	Road Side Units
SDG	Sustainable Development Goals
SuM4All	Sustainable Mobility for All

Funded by the European Union

UC	Use Case
UN	United Nations
VANET	Vehicle ad hoc Networks

1. Introduction

1.1. Project overview

There has been an explosion of mobility services and data sharing in recent years. Building on this, the EU-funded MobiDataLab project aims to create a data sharing culture to let transport authorities, operators, and other mobility stakeholders in Europe know the advantages of sharing data. MobiDataLab develops knowledge and a cloud solution to ease the sharing of data. Specifically, the project continuously creates new knowledge and technical solutions. It collects and analyses the recommendations of experts and supporting cities, regions, clusters, and associations. These actions contribute to the incremental construction of a cross-thematic knowledge base and a cloud-based service platform, which will improve access and usage of data sharing resources.

1.2. Purpose of the deliverable

The objective of this deliverable is to develop an evaluation methodology aimed at assessing: 1) the impact of the data sharing culture and services provided by the new Transport Cloud, 2) the impact of the use cases defined in D2.9 on the environment and the society and 3) the acceptance of the business models by different types of stakeholders. The evaluation methodology is developed based on the analysis of the existing literature, initiatives, and projects. Furthermore, the evaluation methodology will perform surveys during the MobiDataLab events, such as the x-athons. The survey results will be presented in the deliverables D5.2 and D5.3 which implement the evaluation methodology described in this document.

1.3. Intended Audience & Review process

The intended audience of this deliverable is public and private stakeholders who must decide whether to share data or not. The extensive literature analysis and the methodology to assess the impact of the data sharing culture on society and the environment should encourage these stakeholders to be interested in this impact analysis.

Funded by the European Union

1.4. Relation with other work packages/deliverables

This deliverable is related to T3.4 Business and Revenue Models in Data sharing. The objective of the evaluation is to assess the impact and acceptance of the business models. Furthermore, the D3.5 is related to the description of the use cases reported in D2.9, as they are part of the Transport Cloud platform. D3.5 describes the evaluation methodology, while T5.2 Quantification and measurement of the data exchange culture will implement it. The evaluation methodology of the use cases involves the MobiDataLab reference group of stakeholders (municipalities, regions, and others), which is the topic of Task 6.4 (multi-stakeholder group creation and coordination).

2. Evaluation Methodology

The evaluation methodology provided in this deliverable aims to assess the impact of data sharing on society and the environment. Specifically, it provides a framework to evaluate the MobiDataLab use cases and the acceptance of the Transport Cloud and business models.

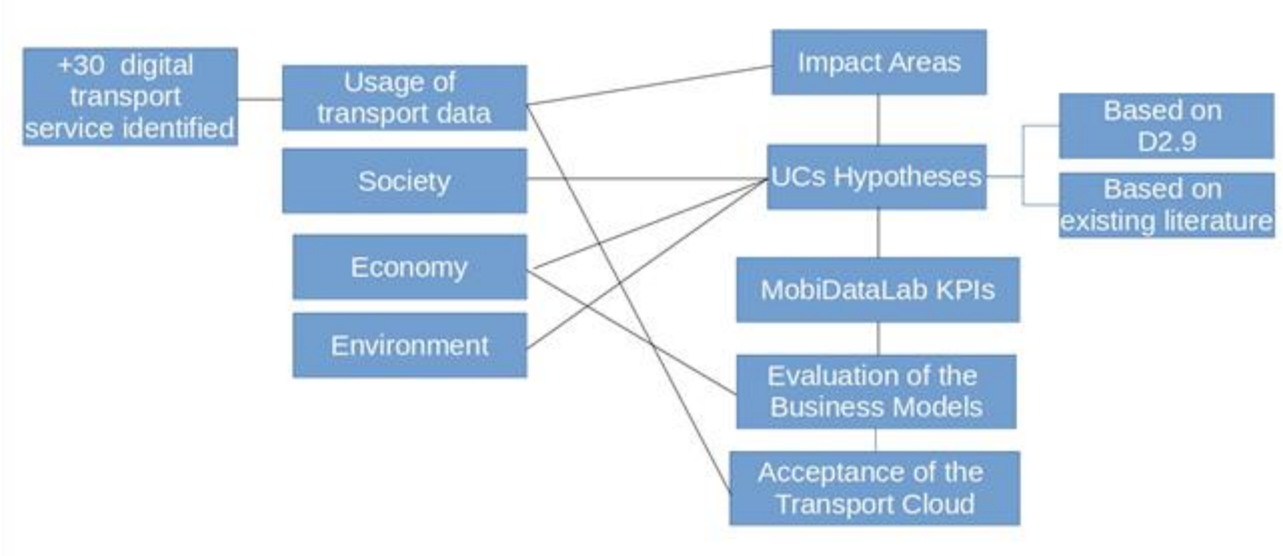

Figure 1: Evaluation methodology

The framework that allows measuring of the mentioned impacts consists of the following steps:

State of the art analysis of the areas of impact of data sharing. The identified impact areas are Usage of Transport Data, Society, Economy, Environment.

The identification of the main study questions to evaluate the impact of the project use cases: a set of study questions will constitute the basis for a survey to be addressed to different project stakeholders to identify how each use case defined in D2.9 may impact society and the environment.

The most relevant KPIs for assessing the impact of the MobiDataLab Transport Cloud: to evaluate the MobiDataLab Transport Cloud impact on different types of stakeholders, projects, and initiatives like MobiDataLab have allowed identifying the main KPIs. Since MobiDatalab will only create a prototype of the Transport Cloud, these KPIs give an idea of how to assess the fully operational Transport Cloud.

The evaluation methodology of the MobiDataLab Transport Cloud and the related business models' acceptance consists of two surveys which will be executed for the entire duration of T5.2, a task that will implement the evaluation methodology.

The first survey concerns the Business Models (BM). It consists of asking questions on the possibility and complexity of implementing each BM in the company the respondent works. The second survey aims to assess the Transport Cloud acceptance, based on the Unified Theory of acceptance and use of technology (UTAUT) and the non-functional requirements of the Virtual Labs, which are virtual services of the Transport Cloud prototype. Both the surveys will be executed by setting up interviews to be carried out during MobiDataLab project events.

3. Data sharing impact areas: State of the art

This section explores and analyses the areas of impact related to the data sharing culture based on the analysis of existing projects related to MobiDataLab and scientific publications. The chapter considers the following data sharing culture areas of impact: 1) the usage of transport data, 2) economy and 3) society and environment. The next subchapters describe in detail each impact area.

3.1. Usage of transport data

Concerning the first area of impact, usage of transport data, a list of more than 30 digital transport services is provided and analysed in terms of their potential impact on society and the environment.

The data sharing culture impacts transport data usage because the more data is shared, the higher the possibility of developing new applications. Innovative digital transport services, such as multimodal travel planners, are based on interoperable, standardised, and integrated data. Demand or supply analysis of transport services and infrastructures deploys transport data. Demand and supply analysis can interest public or private organizations for different reasons, such as understanding the existing gaps in transport services or lack of transport infrastructure. Consultancy companies might take advantage of the data sharing culture to use transport data and perform their studies and analyses. Finally, the research sector might use transport data to develop new methodologies, concepts, or applications. All these stakeholders, especially the public ones, can use data to reduce the impact of transport services on the environment and inform vulnerable people on transport services they might use.

Digital transport services can deploy transport data to create more efficient, safe, reliable, and sustainable transport systems. In this context, sharing data allows integrating different transport modes and enables connected mobility by providing information in real-time. For instance, Mobility as a Service (MaaS) allows connecting all available public transport services of a city, region, or country using a single digital transport system. Connecting public transport services in a digital system is possible when data are open and shared with an open standard format such as the GTFS. According to Sochor et al. (2017), there are four levels of integration for MaaS: 0): zero integration (i.e. travel planning); 1): integration of information (i.e. multimodal travel planner, price info); 2): integration of booking and payment (i.e. find, book and pay for a single trip); 3): integration of the service offer (contracts, discounts and subscription packages across different transport modes); 4): integration of societal goals (MaaS becomes the instrument to achieve the policy goals). Furthermore, Hidalgo (2021) points out that MaaS will allow decarbonisation by connecting suburban and marginalized areas, thus helping to reach societal and policy goals, such as the possibility of living in more sustainable and inclusive cities with more accessible and resilient transport systems.

Digital transport services help create a positive impact on the environment and society. There are studies on specific digital transport services which show why and how they allow reducing the impact on the environment or improving inclusiveness (or reducing the number of accidents). Fagerholt et al. (2010), for instance, point out that fuel consumption and emissions are a cubic function of speed. Therefore, speed optimization between each leg of travel allows considerable savings. This section describes a selection of digital transport services in terms of their efforts to reduce the impact of transport on the environment and create a positive impact on the society by, for instance, fostering inclusiveness of vulnerable travellers or reducing the number of road accidents.

Digital transport services

1. **Vianova** is a data platform that supports cities to better integrate and manage shared, connected, electric and autonomous transport solutions in the urban space, by enabling better use of the infrastructure and promoting safer and more sustainable mobility. Specifically, Vianova is a mobility data platform that facilitates data sharing between cities, operators, and third parties, and it provides analytics and policy tools. It is an open-API platform using open-source data formats and is GDPR compliant. Together with AMAT, the Mobility and Environment Agency of Milan, they developed a shared mobility infrastructure and regulations for mobility. The Vianova platform supports the city of Milan toward its policy of people-centered urban development.

2. **WeWalk** is a tool for visually impaired people that deploys vocal messages and ultrasonic sensors to help them travel by detecting obstacles and accessing public transportation information. Furthermore, the cane allows interacting with the smartphone without removing it from one's pockets. A partnership with Moovit, a MaaS application, has allowed increasing the accessibility of information.

3. **Geotab ignition** captures a vehicle's location, speed, accelerometer data, and other data from the vehicle's computer, including seat belt usage and detailed engine diagnostics. The application allows to manage electric vehicles fleets by monitoring their performances in terms of energy consumption and emissions reporting. In freight delivery, it also provides information on optimal routes, estimated time of arrival (ETA), and allows to assign routes to drivers.

4. **INRIX** aims to manage traffic by analysing data from vehicles. INRIX delivers products for the automotive and transportation industries such as real-time parking and traffic information and solutions that facilitate the safe testing and deployment of autonomous vehicles. The company provides a "green calculator" that allows to reduce the impact on the environment at intersections. Based on an intersection analysis (vehicle volume and current control delay), the Green Signals Calculator calculates the benefits to optimize signals of the network.

Funded by the European Union

5. **Citymapper** is a trip planner which covers all major cities in the world. Together with transit transport, it includes taxi, cycling, scooters and payment possibilities. There is also an API for a routing algorithm and data improved thanks to private feeds. It provides a functionality to quantify the impact of the trip on the environment based on calories burnt, trees and money saved. It also allows to optimize daily commuting trips.

6. **TRAINOTAXI** is offered by the Greek Railway Operator (TRAINOSE) in cooperation with taxi associations in the main Greek cities. It provides affordable taxi transfer from/to the railway station. The combination of train/taxi service has allowed to save about 500 Kg CO_2 per day.

7. **oneTRANSPORT** is a cloud-based platform that allows organisations, including central and local governments and transport service providers, to share and access data about the real-time transport and mobility services. In the UK, four counties with common borders decided to adopt the oneTRANSPORT platform to integrate data into an Azure cloud infrastructure. Thanks to that, several interventions have been possible such as parking optimization schemes, dynamic traffic signal adjustments and improved emergency intervention in case of accidents during high traffic. These have improved traffic flows during peak time.

8. The **OpenDataSoft** platform is a cloud-based data publishing solution that lets users share their data easily. Cities, transportation providers, government administrations, and other private-sector companies use the Opendatasoft platform to publish, visualise, and share data and facilitate their reuse. Thanks to open data sharing there is an increased transparency of the organization. Thanks to data visualization and an interactive dashboard, organizations can communicate their activities and social and environmental impact.

9. **Donkey Republic** is a bike sharing platform aiming to make people use bikes instead of the car, bus or train. The company is committed to sustainable mobility by promoting a new assessment standard which is more transparent and that is recognised by national and multinational institutions that represent the public interest. The social and environmental impacts are obtained due to the increased use of cycling which is emissions free and due to the mental and physical benefits from using bikes.

10. **Carma Carpooling** is an application that allows connecting people and setting up shared trips in real time. The objective of the company is to reduce the number of single passenger car trips. The idea is that leaving three empty seats in the car leads to unnecessary pollution. They also have a partnership with the Climate Protection Campaign, a no-profit organization that aims to create model programs to reduce emissions for communities everywhere.

11. **Glooby** allows to search among millions of flights and hotels to find the most sustainable option. Flights are ranked between 1 and 10, where 10 is the most sustainable solution. Since air travel emits a lot of CO_2 emissions, customers are starting to monitor airlines' sustainable policies. The flight carbon footprint calculator is based on the distance between airports, on the number of stopovers, fuel burn rates (which depends on the aircraft type) and the number of seats per class (first and business class take more space and therefore emit more emissions).

12. **offCents** is an application that aims to reduce CO_2 emissions from transportation by incentivising more sustainable travel alternatives. The approach consists of a market-based approach of carbon offsets. offCents buys and retires carbon offsets from those who decide to participate in the project and then sells them. Buying carbon offsets allows to help fund projects that will reduce carbon emissions equal to the carbon emissions the buyer creates from transportation.

13. **CityCAP** is an application that allows users to earn credits by using environmentally friendly forms of mobility. Every week, a user receives a carbon budget. If the budget is not used because sustainable alternatives are selected, the user receives discounts on services or products. The app allows to collect data on users' carbon footprint, on the time they spend travelling and the distance travelled in different transport modes.

14. **Smou** is an application that can be customized by users to select the services and information that they want to display. The app allows to choose the best way to get according to personal preferences, it shows the location of parking, electric charging points and shared mobility vehicles around the city. Finally, the app predicts the availability of parking spaces and it provides full information on shared bicycles. The possibility to reduce the time people are looking for a parking space allows reducing the CO_2 emissions.

15. **Lyft** is an application and mobility services provider. It offers raid-hailing, vehicles for rent, motorized scooters, bike-sharing, rental cars and food delivery. In their environmental, social and governance report, they claim that they are transforming how consumers access transportation and they commit to have 100% of electric vehicles by 2030. Through LyftUp they connect people with critical resources (foods, jobs, voting) using free or discounted cars, bikes or scooter rides. They have 500 non-profit partners with who they cooperate to offer free rides.

16. **Green Your Move (GYM)** is a multimodal trip planner, funded by the LIFE EU programme, that allows selecting travel routes based on the environmental impact. All types of urban public transport are taken into account. Furthermore, the application considers factors such as weather conditions, air-conditioning use, number of users in the vehicle and the speed of the vehicle.

17. **Nemi** is an application that makes public transport in low-density areas feasible through flexible transport. It allows to optimise the kilometres travelled by public transport means, thus reducing CO_2 emissions and, most importantly, serving people that are usually not reached by traditional public transport.

Mobility as a Service (MaaS)

18. **Whim** is a MaaS that allows combining several transport modes (public transport, city bikes, e-scooters, ferry tickets, taxis, rental cars) into one app. The objective of the company is to replace usage of one million cars, thanks to Whim subscriptions, by 2030.

19. **Meep** is a MaaS application that integrates private and public transport modes into a single platform. It allows to reserve and pay trips and it suggests personalised routes. It aims to improve the accessibility of underutilized transportation systems.

20. **UbiGO** is a MaaS application that combines public transport, car-sharing, rental car services and taxi in one app. It allows to search and book and pay thanks to a monthly subscription. The project has ended in 2021 and participants' countries state that they feel that the environmental impact has reduced thanks to UbiGO.

21. **Moovel** is a MaaS with more than 5 million users which integrates booking and payment services in one application. It is a multimodal mobility platform that allows to integration different mobility providers and that can be used to optimise traffic flow.

22. **ShareNow** is the result of the merger between DriveNow and Car2go, two car-sharing services. Different services will flow from ShareNow: FreeNow for car services with a driver, ReachNow to plan multimodal trips, ParkNow and ChargeNow dedicated to the electric infrastructure. In the future, the fleet will be entirely electric.

Applications for vulnerable users

23. **ENAC PRM** is an application of the Italian Civil Aviation Authority that provides disabled people, people with reduced mobility and elderly practical information on air travel. It provides information on special assistance, i.e., how to ask for it, even when it is not asked in advance or when it is not available or it is inadequate.

24. **Transreport** is an application that allows to better understand the need of passengers. The app also provides passenger assistance during a railway journey. It is necessary to book the assistance in advance to have a railway stress-free travel.

25. **Autism Travel** is a platform that provides useful information to parents of autistic children, with details on the best destinations for such families.

26. **Station Alert UK** is an application for visually impaired people that allows to select a station and set an alert when it is time to get off the train. This is useful as many trains do not announce the train arrival at the station.

27. **Navability** is a travel planner for people with reduced mobility. The route takes into account the gradient, surface and potential obstacles along the route, thus allowing to select the optimal route.

28. **Mapway** is a transit application that provides the London tube map for wheelchair users. It allows people with reduced mobility living in London and tourists to travel the city.

29. **Wheelmap** is an application for people with reduced mobility that provides information on accessible hotels, landmarks and other location in European countries. The application makes traveling easier for disabled people.

30. **SalaBlu+** is an application provided by the Italian Railways, dedicated to passengers with reduced mobility. The app allows you to chat or communicate by phone with the operators that are in charge of assisting people with disabilities. The app also allows to receive information on the trip and there is a recall service.

Applications for the logistic sector

31. **TradeLens** is a blockchain-enabled digital container logistics platform launched by the shipping company Maersk and IBM which uses open standards for data sharing. Thanks to the blockchain, a shared and immutable record of all transactions are kept to secure data in real time.

32. **Onfleet** is a platform that aims to make easy last-mile delivery by providing intuitive routing, dispatching, real-time tracking, analytics, etc. The platform includes end-to-end route planning, dispatch, real-time driver tracking, communication, and analytics.

33. **Uber freight** is an application that aims to better balance demand and supply. The objective is to capture the underutilized capacity to cut freight emissions and reduce fuel costs.

34. **Convoy** is a digital freight network. It provides an application for carriers to find, bid, and haul loads to keep trucks full and earn more. It allows automating matching, pricing, and scheduling to reduce the costs for shippers and allow freights to be delivered safely and timely. This allows to reduce kilometres travelled with empty trucks, thus reducing pollution.

35. **Transfix** is a digital freight platform that aims to reduce emissions through the reduction of empty truck travel. The Environmental, Social, and Governance report, published in 2021 by

Funded by the European Union

Transfix, shows that thanks to the company technologies it has been possible to reduce millions of empty kilometres, preventing thousands of tons of CO_2 emissions.

3.2. Economy

According to the Organisation for Economic Co-operation and Development (OECD), data use and reuse across the economy is a new form of capital. Furthermore, sharing and reusing open data creates new opportunities for economic growth apart from generating benefits for society. The access and sharing of data support economic activities and create benefits for data providers, suppliers, data users, and the overall economy. Several studies, such as from the UK Office of Fair Trading (2016), Deloitte for the UK Department for Business, Innovation and Skills (2013), and ACIL Tasman (2008 and 2009), estimate that data sharing can help to generate social and economic benefits between 0.1% and 1.5% of the gross domestic product (GDP). In the public sector, the GDP growth is between 1% and 2.5% of GDP. An example is the Transport for London (TfL) data reused by several companies that can generate revenues. TfL estimates that data reuse creates between 12-15 million of Gross Value Added per annum, 500 additional new jobs, and 230 indirect jobs.

Therefore, one of the most important characteristics of data to produce economic value is its reusability. However, the possibility of easily finding, accessing, and reusing interoperable data (i.e., FAIR principles) is still remote. According to Springer Nature, researchers face several barriers when using shared data, such as the need to organize them so that they can be useful or the need to understand copyrights issues. For this reason, Springer Nature has published a white paper to accelerate the process of data sharing. In this context, privacy issues are the main concern of stakeholders when sharing data. A review of data protection techniques can be found in Annex II.

The reuse of open data can have different characteristics according to geographical location. When considering the number of organizations that use open and reusable government data, most are developed countries. The Open Data for Development Network (OD4N) is a database of organizations that use open government data. The characteristics of these data are that they are reusable. According to the site, transportation government data are the most used to develop new products and services.

Besides re-usability, there are other relevant characteristics of shared data. The Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitising European Industry (OPENDEI) is an H2020 project that aims to implement the digital transformation in Europe. It aims to create a common data platform based on a unified architecture and standards. Although the project focuses on the manufacturing, agriculture, energy, and healthcare sectors, it is interesting for MobiDataLab because it has identified the fundamental principles for creating data-driven services that enable digital transformation. These principles are:

- Interoperability. Traditionally, ICT services have been developed in a fragmented way. This creates barriers to the development of common data formats and common communication protocols. Interoperability requires the definition of common models and APIs that allow data sharing.

- Openness. All data that can be shared should be open and available either for free or at a reasonable price.

Funded by the European Union

- Reusability. IT solutions, such as open software, should be reusable so that new business models can be developed.
- Avoid vendor lock-in. Digital services and data should be freely transferable among subsystems independently of how and who has developed such subsystems. This principle requires that data should be easily transferred and reused among different applications and systems.
- Security and privacy. To ensure security and privacy for all stakeholders, there should be a common data sharing platform that has agreed on standards, policies, and rules.
- Support to the data economy. This principle is referred to the capability of data to create value. A common data sharing platform will provide value if platforms can connect such that a common marketplace for data is created.

Similarly, Colpaert et al. (2017) present the result of interviews, made with 27 governmental data owners, to identify challenges and solutions to data reuse. The authors start from the consideration that the European Open data portal includes many open transport datasets, but the uptake of these datasets is not high. According to the authors, the reason is the lack of interoperability. Furthermore, interoperability should be evaluated by considering several levels:

- **Legal interoperability.** The open data should comply with the open data definition (opendefinition.org) and have an open license attached to it;
- **Technical interoperability**. It should be easy to bring together two different datasets because there are standardized data formats and the possibility to easily convert from one format to another.
- **Semantic interoperability**. This can be ensured by the Resource Description Framework from the World Wide Web Consortium (3WC) standard which is a general method that allows to describe and exchange graph data.
- **Querying** data. Possibility to querying data.

The Boston Consulting Group points out that data itself does not have value. Data creates value when is transformed into applications and services. Companies often do not recognise the value of data; they can appreciate the value of data if they participate in the data economy.

The value of sharing data in public transportation has been also described by the International Association of Public Transport (UITP) report 2020 Sharing of data in public transport: governance and sustainability. The report underlines the need to create a data sharing culture to make the business models work. Most importantly, the key elements for a healthy data sharing culture are presented. These key elements consist of seeing data as an opportunity but without forgetting the risks of sharing data. There is a need for incentives, findings, and rewards for those who share data and the goals of sharing the data should be transparent. Finally, there should be a responsibility of all the organizations for managing the data which means that not that the responsibility should not be of only one department.

3.3. *Society*

Although businesses have focused on the economic value, data sharing allows for achieving environmental, social, and governance goals. The Boston Consulting Group underlines how companies are pressured by employees, governments, and customers to contribute to addressing social and environmental challenges. They also found that companies that were considered top performers in terms of social and environmental goals, had margins up to 12,4% higher than the other companies.

The report “Enabling data sharing: Emerging principles for transforming urban mobility” of the World Business Council for Sustainable Development (WBCSD) points out that data sharing should be inclusive, ethical, unbiased, and transparent when the objective is to create new policies and outcomes. However, data can be often unbalanced thus leading to discrimination in terms of gender, race, and socioeconomic status. In other words, if some categories are not well represented, discriminatory levels of service could be provided. This can be prevented by setting up representative groups of minorities, ensuring the provenance of data, or by setting up sampling strategies when collecting data such that all of the population is represented. Furthermore, the report underlines the need to create a data governance process that allows for continuously evaluating data based on the feedback from stakeholders.

In line with the report of the WBCSD, the Sustainable Mobility: Policy Making for Data sharing report of 2021, prepared in the context of Sustainable Mobility for All (SuM4All), a coalition of 55 international transport organizations hosted by the World Bank, claims that a key aspect of sustainable mobility is equality for all users. The main problem is that there is limited data availability for vulnerable users (i.e., mobility patterns by genre). On the other hand, the increasing use of digital transport systems allows for obtaining disaggregated data which increases the risks of violating the privacy of users. In this context, policymakers should protect vulnerable travellers.

Concerning social inclusion, the consultancy company Bax & Company points out that it is necessary to foster the development of new digital mobility services to achieve sustainable mobility. In this context, the H2020’s HiReach and INCLUSION projects have identified the key characteristics of digital transport services that can foster inclusivity for the vulnerable, such as the elderly and citizens with disabilities. These key characteristics of digital transport systems are the following:

1. Availability of transport: digital transport services should include a great variety of routes and services and information on delays, and transport services on-demand while the information should be provided in multiple languages.

2. Accessibility: innovative digital transport services should allow the vulnerable to access services, for instance, by providing new functionalities such as voice recognition. The digital transport services should also give information on physical accessibility and the best mobility solution.

3. Affordability: not all transportation is necessarily affordable. Personalized offers for vulnerable users should be publicised. An example is the Washington Metropolitan Area Transit Authority which offers discounts to late-night commuters taking a Lyft shared ride.

4. Adequacy: is a combination of factors that are important for the user. These factors can be comfortability, reliability, safety, cleanness, easiness to use.

Funded by the European Union

5. Awareness: Users should be aware of the available offers and their characteristics. In this context, local authorities should promote the use of digital transport services by training users and creating guidelines.

Lucas (2012) has shown that multiple factors contribute to social inclusion. Personal characteristics, geography, institutional, mobility, and accessibility factors can either reduce social exclusion or increase it. Stanley et al. (2021) found that an additional trip made by socially excluded people has a high value.

The characteristics of inclusive digital transport services evidence the need for a mature data sharing culture to have inclusive digital transport services.

3.4. *Environment*

The reduction of the impact on the environment due to the data sharing culture and digital transportation is explained by Datumize, a software company for data integration and digital transformation. More specifically, the company explains how digital transformation in logistics affects the industry which contributes to the13% of emissions. Digital solutions allow the implementation of new business models such as green and reverse logistics which apply the principles of circular economy to logistics. Through digital solutions, these business models are easier to implement by providing, for instance, automated inventory solutions. The platforms such as Uber Freight, Convoy, or Transfix allow to lower the costs, improve the efficiency of the logistic chain and reduce the environmental impact. Furthermore, telematics applications also allow for reducing the impact on the environment, and the Internet of Things (IoT), robotics, and big data analytics allow for better collaboration. IoT allows to track and trace thus optimizing flows and improving the efficiency of the transportation chain.

The Federated Network of Platforms is an EU project that aims to create a federated network of platforms for data sharing in the logistics sector. The project promotes the need to establish federated agreements on data sharing to achieve energy efficiency throughout the supply chain. The possibility to make informed decisions on the energy used for transport, which is possible due to data sharing, is the basis for green transport in the multimodal chain. According to Lind et al. (2021), energy efficiency in the transport sector can be achieved by involving actors and by making them able to make informed decisions on energy use. This allows to synchronize actors and have multimodal green transport. However, having only a few local digital transport systems is not enough to enable green conversion. The authors claim that agreements on which data, when, and how to share data need to be established in the context of federated platforms. They also report some examples of applications of digital data sharing for green conversion:

- Carbon footprint recorder. There is a limited measure of the CO2 emissions from the supply chain, and this is due to the high fragmentation, lack of interoperability, and information sharing.
- Matching demand and supply of fossil-free fuel. The use of green energy and alternative fuels by transportation means is a prerequisite to protecting the environment. In this context, there is a need for digital data sharing and platforms to match the demand and supply of these fuel stations.

- Smart grids and load balancing.
- Matching supply of available transport capacity needs.
- Situational awareness to re-direct transports based on sequential decision making.
- Marketplace for available components to be recycled. The possibility to recycle materials from used vehicles is part of the circular economy and must be incentivized through the establishment of digital marketplaces.
- Green approaches to chains of transport nodes. The key to greener transport is multimodal digital data sharing which will enable the development of new digital services.

Concrete examples of how digital transport systems and data sharing can have an impact on the environment are reported in Chapter 3.1 in which a list of digital transport services is reported and described in terms of their capability of having an impact on the environment and/or on the society.

4. Hypotheses and study questions to assess the social and environmental impact of the use cases

This chapter aims to formulate a set of hypotheses based on the description of the Use Cases (UC) reported in D2.9. Based on the hypotheses, a set of study questions are formulated to identify the impact of the Use Cases and of sharing data on society and the environment. Based on the proposed study questions, a survey will be developed in T5.2:

- The first analysis will be performed based on the survey addressed to the project participants. Results will be provided in D5.2.
- Later, the survey will be extended and performed with respondents who are not part of the project and belong to different categories of stakeholders. In this latter case, the analysis of the interviews will be presented in D5.3.

The analysis and investigation have been done for UCs 1-6. The UCs 7-9 refer to specific research topics for which their social and/or environmental impact does not worth an additional investigation, due to the fact that the impact is implicitly defined in the specified objectives (UC7 will have an impact on people with reduced mobility; UC8 on environment, UC9 on tourism stakeholders).

UC1 (Optimization of transport flows and ETA)

The UC1 (Optimization of transport flows and ETA) aims to deploy data that allow for predicting the Estimated Time of Arrival (ETA). The optimization of transport flow can be static when the route is pre-planned or dynamic when the driver can adjust the route thus influencing the ETA. According to D2.9, the dataset for ETA can be static or dynamic which means that traffic data can be in real-time or based on static map data. Additional data can be considered such as weather conditions, rest

Funded by the European Union

time regulations, and planned events. Concerning operational data, there is a need for telematics data of the vehicles, their location, completed stops, tour plans, and driver shift time.

Table 1: UC1 Optimization of transport flows and ETA: Hypothesis and Questions

Hypothesis	Questions
The optimization of transport flows with an estimated time of arrival can reduce the impact of freight transport on the environment.	How can optimization for freight deliveries with an estimated time of arrival reduce the impact of transport on the environment?
The optimization of transport flows with an estimated time of arrival can improve the social conditions of citizens	In which way, in your opinion, the route optimization for freight deliveries with estimated time of arrival can improve the social conditions of citizens?
The optimization of transport flows with an estimated time of arrival (ETA) can create a positive impact on society.	How, in your opinion, the optimization of transport flows with ETA can create a positive impact on the quality of work of the freight delivery of company employees?
The optimization of transport flows with an estimated time of arrival (ETA) can inform policymakers.	How can optimization of transport flows with ETA inform policymakers?
Some critical factors that allow for an impactful ETA implementation.	In each of these, what is the most critical factor for an impactful ETA implementation? (Only, if necessary, mention alerts for delayed stops, sharing arrival time, post-trip reporting, update tour plan, rest time planning)

Several indicators can be considered to evaluate the impact of the Optimization of Transport Flows and ETA. The definition of these indicators depends on the type of transport mode. For the logistics services, it might be important to increase the number of on-time deliveries. The optimization of transport flows is important for all transport modes because it allows reducing the cost of transport per unit thanks to the decrease of the km travelled and, consequently, the cost of fuel. Based on these considerations, Table 1 reports a set of hypotheses to be tested in the context of the Living Labs demonstrations.

UC2 (Emissions Reporting)

Concerning the UC2 (Emissions Reporting), there are several works in the literature that aim to assess whether the emissions reporting has an impact on environmental sustainability. According to the *Proving the Case: Carbon Reporting in Travel and Tourism*, a report written by Amadeus, a leading travel technology company, and the University of Griffith, there is an increasing number of travel and tourism companies that are engaged with emissions reporting. Some tools assist companies in developing initiatives that allow them to reduce their impact on the environment.

Overall, there are several tools available but there is often a lack of transparency. At a company level, there are two main frameworks for emissions reporting: The Carbon Disclosure Project (CDP) and the Global Reporting Initiative (GRI). Within the framework, there are the emissions calculators and the corresponding indicators. The CDP was established in 2000 as an independent organization that distributes information on emissions. It is based on the principle of "measure to manage" and the idea is to provide incentives for companies or cities that reduce their impact. The GRI is a non-governmental organization that aims to drive emissions reporting to reach environmental sustainability goals. The authors point out that academia is more focused on finding indicators that measure the cost of travel in terms of carbon emissions; while the industry is more focused on how to reduce emissions. Overall, the existing literature on the capability of the emissions reporting to reduce the impact of transport or economic activities on the environment is missing. On the other hand, some works propose innovative ways of measuring environmental impact, but the main problem is transparency in data sharing. Based on these considerations, a set of hypotheses and the corresponding questions aimed at understanding how the emission reporting can help to reduce the impact on the environment and society is reported in Table 2.

Table 2: UC2 Emission reporting: Hypothesis and questions

Hypothesis	Questions/Interviews
The emissions reporting allows for reducing the impact of transport on the environment	How, in your opinion, can the emission reporting help to reduce the impact of transport on the environment?
The emission reporting produces a positive impact on the society	How can the emission reporting produce a positive impact on society?
The emission reporting is able to create a positive impact on society.	How the emission reporting can produce a positive impact on the society?
Emission reporting is complex and has several critical aspects.	We know that emission reporting is a complex business. What are the main critical aspects of emission reporting, according to you? (reporting indirect and direct emissions, attributing emissions?)
Once the challenges related to the emission reporting are addressed, it will be possible to maximise the social and environmental impacts.	How can we overcome the challenges of emission reporting to maximize social impact? What about environmental impact?
Stakeholders already have good practices regarding the emission reporting.	To which extent do you think stakeholders in this project have already good practices in terms of report emissions, the report predicted emissions, and history of past emissions?
There is willingness and commitment to develop common standards and practices.	To which extent do you think there is willingness and commitment in the network to develop common standards and practices for report emissions?
There is enough transparency and honesty in emission reporting.	To which extent do you think there is transparency and honesty in report emissions?

UC3 (Analytics and Learning)

The UC3 (Analytics and Learning) is very important regarding the development of a data sharing culture. The development of innovative approaches that can help to reduce the impact of transport on the environment depends, in fact, on the availability of data. In Castaneda et al. (2021), for instance, a machine learning model is developed to predict the level of sustainability of freight transport companies based on a set of KPIs. Their objective is to allow companies to find strategies to reduce the impact on the environment. The possible applications of machine learning to the transport sector are many. Machine learning and data analytics can be deployed to understand and predict the behaviour of travellers thus suggesting to them the best solution for their trip. Machine learning can be deployed to learn from other impaired users which are the best solutions in terms of routes and transport modes and provide suggestions to travellers with similar characteristics. In this case, the availability of data will be of fundamental importance and the risk of training the algorithm using data that under-represents some categories will be high. Therefore, the objective of the hypothesis and corresponding questions reported in Table 3 is to understand how these techniques can have an impact on the social and environmental dimensions.

Table 3: UC3 Analytics and Learning: Hypothesis and questions

Hypothesis	Questions/Interviews
Data analytics helps to increase data quality, data sharing and collaboration.	What do you think are the key factors for sharing high quality data? In particular, 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?
Data analytics and learning help to reduce the impact of transport on the environment.	How can data analytics and learning techniques help to reduce the impact of transport on the environment, in your view?
Data analytics and learning help to create more inclusive digital transport services.	Do you think that data analytics and learning can be of help to create more inclusive digital transport services?
Data analytics and learning increase the ability to interpret and use data for decision making.	What about the ability to interpret and use data for decision making?
Data analytics and learning help to increase experts' knowledge in the mobility sector. This will have a positive impact on the society and the on employment.	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?

UC4 (Reuse of transport data for journey planners / digital transport service)

The UC4 (Reuse of transport data for journey planners / digital transport service) clearly shows the importance of sharing data to integrate different transport modes, thus, making multimodality possible. This UC is mainly considering passenger transport travels. However, in the future, thanks to a broader introduction of IoT and shared transportation, it will be possible to channel freights through a multitude of transport modes such as rail, bike, and electric mobility.

According to Cheung and Sengupta (2016), data quality is the most important factor of a journey planner. For this reason, shared data for multimodal journey planners should be evaluated in terms of the completeness, consistency, accuracy and integrity of the dataset which also determines the reliability of a journey planner. The main issue of journey planners, thus, is the possibility to access multiple data sources to enable multimodality. In Colpaert et al. (2017), a Federated Route Planning is proposed as the multimodal route planning that provides route planning advice and access to multiple datasets. According to Poom et al. (2020), journey planners promote healthier and more sustainable urban travel. In Zuurbier et al. (2010), the authors point out that exposure to pollution depends on travel mode. Cyclists and pedestrians are exposed directly to emissions compared to public transport and private transport travellers. Therefore, the routing algorithm should not only consider the travel time or the distance but it should also take into account environmental information. This allows suggesting more pleasant and healthier route alternatives to vulnerable to pollution exposition travellers. From the point of view of the impact on society, the journey planner can suggest safer alternatives, based on the availability of sidewalks, street width, cycleways or routes that are more pleasant in terms of aesthetics. Based on these considerations, Table 4 reports the hypothesis of impact on the environment and on the society and the corresponding questions.

Table 4: UC4 Re-use of transport data for journey planners / digital services: Hypothesis and questions

Hypothesis	Questions/Interviews
The reuse of transport data for journey planners can prevent the exposure of travellers to environmental pollution.	How, in your opinion, could the journey planner create positive social and environmental impact?

UC5 (Mobility as a Service / MaaS)

The UC5 (Mobility as a Service / MaaS) partially includes the journey planner which has been previously discussed. The MaaS is an "integration of various forms of transport services into a single mobility service accessible on-demand" (*White Paper, MaaS Alliance, 2017)*. Three levels of MaaS implementation are described by Koichi (2020). In Level 1, MaaS only provides integrated information and revenues come from advertisement; Level 2 MaaS includes payment and revenues of the MaaS provider can be derived from usage fees and payment to the service provider; finally, in the case of Level 3 MaaS, users can pay a fixed amount for the service usage.

MaaS is expected to change travel behaviour by creating new opportunities for public transportation (Shaheen et al. 2020) and it has a great potential to satisfy the mobility needs in a sustainable way (Giesecke et al. 2016). The MaaS Market Playbook of the MaaS Alliance affirms that public authorities should establish a framework to ensure compliance with societal and environmental goals. Furthermore, the presentation held by Arthur D. Little, at the MaaS Alliance Plenary meeting on 5 October 2020 shows some of the trends post COVID which could accelerate the realisation of

Funded by the European Union

MaaS. In terms of behavioural trends, there is more consciousness of travel safety, healthier mobility lifestyle and evolution of trip patterns. In terms of technological trends, there is an increased offer of technologies, intelligent transport systems and more availability to accept new forms of mobility.

The MaaS Market Playbook points out that the capability of MaaS to offer a combination of environmentally friendly mobility services allows travellers to contribute to improving the environment and the quality of life. Most importantly, the report concludes stakeholders should share their experiences and results because politicians need more data about the benefits and impact of MaaS to make the society less polluted, less congested and more liveable. Based on these considerations, Table 5 reports the hypothesis and the corresponding questions to determine which are the impacts of MaaS on society and the environment.

Table 5: UC5 Mobility as a Service (MaaS): Hypothesis and questions

Hypothesis	Questions/Interviews
MaaS is able to influence the behaviour of travellers who will find it easier to access and use a mix of more sustainable mobility services.	How does MaaS contribute to foster multimodal mobility and the use of new forms of mobility that allow to reduce the impact of transport on the environment? (Less congestion, pollution?)
MaaS is able to improve the liveability of cities.	How can MaaS improve the liveability of cities?
Exogenous events foster greater attention to MaaS and journey planning tools.	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?
Exogenous events increase social attention and relevance of the project.	What could be some ways to increase social attention and relevance of the project at the light of these social shocks?

Main findings

The assessment of the impact of the Use Cases on society and the environment is a complex task because there are multiple ways the Use Cases can be implemented. This means that, according to the type of implementation, the Use Case could have different effects. If the journey planner includes a functionality that allows visually impaired people to access public transport, these vulnerable people will benefit the most. However, the intensity of the impact also depends on how many users take advantage of the service and to what extent the service is useful. The usefulness of the digital transport service also depends on how much data is used to provide information to the end-users. This means that the same digital transport service might be useful in one city but not in another because data is missing. Here comes the data sharing culture that will allow implementation of digital transport services in different cities and locations of the world and data integration to enrich information. The exploration of the impact of the MobiDataLab Use Cases will continue and will deepen after the interviews with project participants will be carried out.

5. Assessing the potential impact of the MobiDataLab Transport Cloud: lessons learnt and KPIs from other projects and initiatives

The majority of projects and initiatives on open data and data sharing deal with the provision of open data for research purposes. However, some of the KPIs and evaluation metrics that have been identified by these projects, might be relevant for MobiDataLab. Therefore, the purpose of this chapter is to analyse existing projects, studies and initiatives to identify indicators and adapt those indicators to measure the impact of the MobiDataLab Transport Cloud. Since the Transport Cloud developed by MobiDataLab will only be a prototype, it will not be possible to concretely measure the majority of the indicators. However, these indicators are interesting because they can measure the acceptance of the Transport Cloud by different types of stakeholders. For each of the selected projects, a subset of KPIs will be considered and readapted to MobiDataLab. Finally, at the end of the chapter, there will be a summary table in which, each of the readapted KPIs will be associated with a specific stakeholder.

The European Open Science Cloud (EOSC) (eosc-portal.eu) is a federated and multidisciplinary environment that allows the publication of FAIR data. The prototype components of the platform are developed from 2021 to 2030 through the Horizon Europe funding. BLUE CLOUD, for instance, is an H2020 federated platform of marine data infrastructure. It allows the combination, reuse, and sharing of data across disciplines and countries. The success of the EOSC is measured in terms of the uptake by researchers. The EOSC Secretary (2020) has indicated a list of KPIs. Table 6 reports a selection of the KPIs adapted to the MobiDataLab Transport Cloud.

Table 6: MobiDataLab KPIs based on EOSC Secretary (2020)

EOSC Secretary (2020) KPIs	**MobiDataLab Transport Cloud KPIs**
Researchers performing publicly funded research make relevant results available, as openly as possible.	Public organizations and publicly-funded projects make available and share datasets as openly as possible, through the Transport Cloud.
Research data produced by publicly funded research in Europe is FAIR by design.	Data produced by publicly funded research projects in Europe and shared through the Transport Cloud is FAIR.
The EOSC Interoperability Framework supports a wide range of FAIR digital objects including data, software and other research artefacts.	The MobiDataLab Transport Cloud supports a wide range of FAIR digital objects including data, software and other research artefacts.
European research is increasingly discovered and reused across disciplines as a result of EOSC.	Transport data is increasingly discovered and reused across disciplines as a result of the Transport Cloud.

Funded by the European Union

EOSC is operational and provides a stable infrastructure, supporting researchers in addressing societal challenges.	The Transport Cloud is operational and provides a stable infrastructure, supporting public and private organizations in addressing societal challenges.
EOSC is populated with a valuable corpus of interoperable data.	The Transport Cloud is populated with a valuable corpus of interoperable data.
EOSC is a valuable resource to a wide range of users from the public and private sectors	The Transport Cloud is a valuable resource to a wide range of users from the public and private sectors.

OpenAIRE-Connect - CONNECTing scientific results in support of Open Science is an H2020 project that aims to provide open access to research findings. The project participants contribute to the EOSC by taking part in working groups and supporting the coordination of the EOSC portal.

Table 7: MobiDataLab KPIs based on OpenAIRE

***OpenAIRE KPIs on the Portal* KPIs**	**MobiDataLab Transport Cloud KPIs**
Number of H2020 datasets	Number of datasets made available through the Transport Cloud.
Number of H2020 Open Access datasets	Number of open datasets made available through the Transport Cloud.
Number of data providers	Number of companies / institutions that share data through the Transport Cloud.
Number of registrations by country	Number of registrations by country.
Number of registered end-users of the portal	Number of registered end-users of the portal.

BE OPEN - European forum and oBsErvatory for OPEN science in transport, beopen-project.eu, is an H2020 Coordination and Support Action that aims to promote Open Science in transport research. The project aims to set up the Transport Observatory / fOrum for Promoting Open Science (TOPOS). In D5.2: KPIs for Open Science evaluation in transport, a list of KPIs is identified to measure and monitor the implementation of open science in the transport sector. More specifically, a set of categories and corresponding KPIs have been identified to measure the success of TOPOS and Open Science in transport research. Table 8 reports a selection of KPIs that have been adapted to the MobiDataLab Transport Cloud.

Table 8: MobiDataLab KPIs based on BE OPEN D5.2 KPIs for Open Science evaluation in transport

Domains	**MobiDataLab Transport Cloud KPIs**
Data sharing	Number of datasets being published in the Transport Cloud

	Number of organizations / institutions that take part in the Transport Cloud Number of datasets by association/organization % data FAIR compliant
Exchange knowledge	Number of views and citations of MobiDataLab Transport Cloud content
Validation of data	Number of publicly funded openly available datasets
Resources and organizational issues	Number of organizations with a process in place to share open data thanks to the MobiDataLab Transport Cloud
FAIR metrics and certification	% Organizations that have FAIR dataset deposited in the Transport Cloud % Datasets deposited in the Transport Cloud that are FAIR
Ensure Data Ownership and IPR	Number of datasets in the Transport Cloud which are compatible with the EOSC licensing model

In the context of the **OpenDataMonitor** project, Atz (2014) proposes a new metric to assess the timeliness of data in catalogues. Timeliness is important because business and start-ups want to be sure that the data that they use for their business remain available and updated. The author proposes the **tau** measure which is the *percentage of datasets up-to-date in a data catalogue*. The author defines timeliness:

$$timeliness = I(\frac{updatefrequency}{today - lastsubstantialupdate})$$

The indicator is 0 or 1 and it depends on the last substantial update. While *I* is an indicator function that takes the value of 1 if the ratio is bigger than 1, 0 otherwise.

$$tau = \frac{1}{N}\sum_{i=1}^{N} I(\frac{updatefrequency_i * lambda + delta}{today - lastsubstantialupdate_i})$$

Where N is the number of the dataset available in the catalogue, delta is the fixed number of days the catalogue is allowed for updating (i.e. one day for processing). Lambda is relative to the update frequency. A tau of 0 means that the catalogue has no update on the dataset. A tau of 1 means that all dataset is updated. Where update means a substantial update of the dataset.

Additional indicators of the data catalogue could look at the timestamp of data and look at the **data ranges** the dataset covers. Furthermore, it is interesting to see how the tau changes over time and based on the geographical location of data or the type of sectors it covers.

Funded by the European Union

Table 9: Summary of MobiDataLab KPIs

Source	MobiDataLab KPIs to measure the acceptance of the Transport Cloud	Type of stakeholder
EOSC Secretary (2020) / BE OPEN	Data produced by publicly funded research projects in Europe and shared through the Transport Cloud is FAIR.	Public and private organizations, Research institutions
EOSC Secretary (2020)	The MobiDataLab Transport Cloud supports a wide range of FAIR digital objects including data, software and other research artefacts.	Public and private organizations
EOSC Secretary (2020)	Transport data is increasingly discovered and reused across disciplines as a result of the Transport Cloud.	Public and private organizations
EOSC Secretary (2020)	The Transport Cloud is operational and provides a stable infrastructure, supporting public and private organizations in addressing societal challenges.	Governments, public and private organizations
EOSC Secretary (2020)	The Transport Cloud is populated with a valuable corpus of interoperable data.	Public and private organizations
EOSC Secretary (2020)	The Transport Cloud is a valuable resource to a wide range of users from the public and private sectors.	Public and private organizations
OpenAIRE/ BE OPEN	Number of datasets made available through the Transport Cloud.	Public and private organizations
OpenAIRE / BE OPEN	Number of open datasets made available through the Transport Cloud.	Public organizations
OpenAIRE / BE OPEN	Number of companies / institutions that share data through the Transport Cloud.	Public and private organizations
OpenAIRE	Number of registrations by country.	End users, public and private organizations
OpenAIRE	Number of registered end-users of the portal.	End users, public and private organizations
BE OPEN	Number of datasets by association / organization.	Public and private organizations
BE OPEN	% of datasets deposited in the Transport Cloud that are FAIR.	Public and private organizations
BE OPEN	Number of uploads in the context of projects granted by the EC	Public and private organizations
BE OPEN	Number of views and citations of MobiDataLab Transport Cloud content	Public and private organizations
BE OPEN	Number of organizations with a process in place to share open data thanks to the MobiDataLab Transport Cloud.	Public and private organizations

6. Acceptance of the Business Models and the Transport Cloud

This chapter aims to present the framework to evaluate the Business Models defined in D3.4 and the acceptance of the Transport Cloud. Qualitative surveys will be carried out during x-athons and other MobiDataLab events. There will be two separate surveys: one for the evaluation of the Business Models and one to assess the acceptance of the Transport Cloud. The objective is to interview different types of stakeholders such as digital transport service providers, public organizations, researchers, innovators, private companies, start-ups and end-users such as cities that might use the Transport Cloud. If it will not be possible to interview one or more categories of stakeholders, additional interviews will be organized with external respondents.

6.1. Evaluation of the MobiDataLab Business Models

Lessons learnt from the literature

The literature provides abundant templates and processes for developing BMs (among the others: Ehret et al., 2013; Joyce and Paquin, 2016; Gordijn and Akkermans, 2001; Hayes, Finnegan,2005). However, nowadays, most of the BMs are represented using CANVAS which allows to highly standardise the components of the BM (Lizarralde et al, 2020).

The BM Canvas (Osterwalder and Pigneur, 2010) was originally introduced by Alexander Osterwalder starting from his previous work regarding the BM ontologies (Osterwalder, 2004). According to the author, a BM should describe how an organization creates, delivers and captures value and, in order to have all the information available in a single diagram, the BM CANVAS template was created. Even if it is the most used template, this representation of the BM can be difficult to be understood by the stakeholders who need to clearly figure out their possible future impacts (Gordijn and Akkermans, 2001). In other words, the BMs represented using the CANVAS models are static and, therefore, it is not possible to understand their predicted impact or their strengths. Furthermore, this representation of the BM can be difficult for the stakeholders to understand, especially if they need to assess if one BM is more effective compared to another.

The evaluation of the business models is a complex activity as it requires close collaboration between many people (diversified by age, experience, role) who can combine their different points of view in order to define a BM that is well structured and concrete (Osterwalder and Pigneur, 2010).

As for the evaluation of the BM, according to Haaker et al. (2017), the main focus when evaluating a BM is robustness. Several aspects might influence the robustness of business models such as the unpredictable change in technologies, regulations and evolution of the market. To be sure that the BM is solid, the authors propose a methodology to assess the robustness of the business models. The approach consists of six steps:

- Description of the business models using CANVAS, STOF, VISOR or MB Cube. This is the first starting point for evaluating BMs and, the advantage of the model proposed by Haaker et al. is precisely the possibility of starting from any template used to develop the BM.

Funded by the European Union

- Identify the stress factors as trends, uncertainties and outcomes (derived from existing scenarios or brainstorm sessions with stakeholders);
- Explore the stress factors and find which business model components are casually related to them;
- Evaluate how the stress factors affect the business models through a heat map (a matrix with business models positioned vertically and outcomes on uncertainties horizontally). The heat map is characterized by a colour scheme that helps indicate the impact of a specific stress factor on a business model component.
- Analyse the results starting from the Heat Map that will show which business model components are robust.
- Formulate improvements and actions. This final step aims to define actionable conclusions and recommendations on how to improve weak business models components or on how to improve consistency.

Gordijn and Akkermans (2001) propose the e^3 value modelling approach to e-business. The authors point out that business models are usually characterized by a mix of text and graphics but that these representations are not always easy to understand, especially for the stakeholders. Therefore, they proposed a new way to describe BM: an e-business modelling approach that is able to combine the IT system analysis and the economic value perspective from business science. They also underline the importance of evaluating the economic feasibility of e-business models: the idea – according to the authors, needs to be analysed in quantitative terms, always considering that, to be feasible, an e-business model must give a profit to all actors involved. Based on these considerations, the evaluation of e-business models takes into account the net in and outflows of values objects. More specifically, the difference between inflows and outflows should be sufficient to cover all expenses. Furthermore, the authors propose a what-if scenario analysis concerning financial parameters, future trends and customer behaviour.

Regarding the evaluation of the BMs, Gilsing et al. (2021) propose an approach to defining the Business Models KPIs to support decision making. Their approach is based on the linguistic summarization theory which is a fuzzy set theory that aims to take into account uncertainty. In a fuzzy set, a pair of elements belong to the set with a certain degree of membership. The authors make the example with the age of a car that can be new, used or old where each of the linguistic values can be represented with a fuzzy set. Linguistic summarization are natural language sentences which consist of protoforms (e.g. most cars are new, most new cars are fast). Through their approach, they are able to identify soft KPIs that, in the authors' opinion, are the most useful in the early innovation process.

Another interesting point of view on how to evaluate business models is the one proposed by Patrik Budsky (Budsky, p. 2019). He presents the process of calculating the added value of a business model by valuing and comparing different scenarios with different business models. The process proposed by Busdky foresees the usage of a calculation tool as Free Cash Flow or Enterprise Value Added. The evaluation methodology presented in this paper, however, can only be applied when a milder innovation of the business model is found.

Funded by the European Union

The MobiDataLab approach

In MobiDataLab the main interest is to test the acceptance of the BMs by the potential users. By summarising the lessons learnt from the literature analysis, Table 10 reports the list of selected questions to evaluate the MobiDataLab Business Models.

Table 10: Questions to assess the acceptance of the Business Models by the user

Questions	Measure
Is this BM attractive for our organisation?	Five-point Likert scale (not at all attractive–very attractive)
Is this BM feasible for our organisation?	Five-point Likert scale (not at all feasible–easily feasible)
How scalable is this BM?	Five-point Likert scale (not at all scalable–very scalable)
How costly is the set of this BM?	Five-point Likert scale (very cheap–very costly)
How inclusive is this BM?	Five-point Likert scale (not at all inclusive–very inclusive)
How attractive is this BM for private entities?	Five-point Likert scale (not at all attractive–very attractive)
How attractive is this BM for public entities?	Five-point Likert scale (not at all attractive–very attractive)
How technically complex is the BM?	Five-point Likert scale (not at all complex–very complex)

The questionnaire reported in Table 10 will be answered in the context of interviews organized during MobiDataLab projects events. Private companies and start-ups will be the most interesting respondents. Therefore, the efforts should be to get feedback on the Business Models from these two types of actors. While public organizations are usually more interested in cooperation models. A description of the BM is provided in D3.4.

Funded by the European Union

6.2. Acceptance of the Transport Cloud

The Unified Theory of Acceptance and Use of Technology (UTAUT) was selected to assess the acceptance of the Transport Cloud. The UTAUT has already been deployed to assess the acceptance and use of open data technologies. Furthermore, the survey created based on the UTAUT has been enriched using the questions to calculate the Global Open Data Index and based on the questions derived based on the MobiDataLab Virtual Labs non-functional requirements.

The UTAUT has been proposed by Venkatesh et al. (2003) who analysed several models of acceptance that studied how and why people use Information Technology and tried to unify them. Through the comparison of eight models, they found commonalities that directly impacted user acceptance and user behaviour: *performance expectancy*, *social influence* and *facilitating conditions*. Performance expectancy is related to the belief that the technology will help the user to facilitate the job and increase performance. The authors found that performance expectancy is the most important factor for predicting intention to use the technology. Furthermore, according to the authors, this variable is strictly related to gender because men are usually more task-oriented although the culture might have a stronger influence. Effort expectancy is related to the easiness of using the technology. In this case, gender, age and work experience are considered to be influencing factors. Social influence refers to the fact that individuals consider important to use technology that is accepted by society (e.g. using Facebook because it is socially acceptable). Finally, facilitating conditions refer to the belief that there is enough support from the organization to use the system. Using structural equations, the authors show that their model is able to explain the 70% of intention to use technology.

The acceptance of open data technologies has been analysed by Zuiderwijk et al. (2015). The authors identified the factors that contribute to the acceptance of the open data technologies which can be taken into account to stimulate the use of open data. Their work is based on the UTAUT theory with some modifications of the hypotheses which are adapted to the open data framework.

Table 11: Dimensions of intention to use and acceptance of open data technologies (based on Zuidderwijk et al., 2015)

UTAUT construct	Explanation	Correlation with intention to use and accept open data technologies
Performance expectancy (PE)	The use of open data technologies (platforms, software, tools and interfaces) is expected to help the worker to perform better.	positive
Effort expectancy (EE)	The heterogeneity of open datasets, lack of interoperability, diversity, and fragmentation can create barriers when using open data. For this reason, the perceived difficulty of using open data could translate into increased effort expectancy.	negative
Social influence (SI)	Promotion of using open data by the social environment could incentive the willingness to use and accept them.	positive

Facilitating conditions (FC)	For some people access to the internet is not granted and, in this case, access to open data is also difficult.	positive
Behavioural intention (BI)	The intention of an individual to use an open data technology in the future.	positive
Voluntariness of use (VU)	The use of open data may be required by the company. In this case, the use of open data is not voluntary. Therefore, the intention to use and the acceptance of open data are negatively correlated with the voluntariness of using.	negative

Based on the variables and the questionnaire developed by Zuiderwijk et al. (2015) to assess the intention to use and accept open data technologies, a questionnaire is created to evaluate the intention to use and acceptance of the Transport Cloud. The questions have been adapted based on the characteristics of open data which, thanks to the Transport Cloud, will become FAIR.

The survey will be executed during the hackathons organized in the context of MobiDataLab and during other types of project's events. The questionnaire has to be filled out after having explained what the Transport Cloud is and its foreseen functionalities. Furthermore, the Social Influence UTAUT construct is not considered in this analysis because the Transport Cloud is a prototype. The Hypotheses of correlation reported in Table 14 will be tested and principal component analysis and multiple regression will be implemented to assess the percentage of variance that the variables listed in Table 12 are able to explain in terms of the intention to use and acceptance of the Transport Cloud.

Table 12: Questionnaire to assess the intention to use and accept the Transport Cloud (based on Zuiderwijk et al., 2015)

UTAUT construct	Question	Type of outcome
Performance expectancy (PE)	Using the Transport Cloud is of benefit to me (PE1) Using the Transport Cloud will enable me to accomplish my research more quickly (PE2) The possibility of easily finding mobility data thanks to the Transport Cloud will increase my productivity (PE3) The possibility of easily finding open data thanks to the transport Cloud will increase my performance in my job (PE4)	Five-point Likert scale (strongly disagree–strongly agree)
Effort expectancy (EE)	It will be easy for me to become skilful at using the Transport Cloud (EE1) Learning to use the Transport Cloud will be easy for me (EE2)	Five-point Likert scale (strongly disagree–strongly agree)

Funded by the European Union

	I clearly understand how to use the Transport Cloud (EE3) I do not have difficulties in explaining why using the Transport Cloud may be beneficial (EE4)	
Social influence (SI)	None	Five-point Likert scale (strongly disagree–strongly agree)
Facilitating conditions (FC)	I have the resources necessary to use the Transport Cloud (FC1) The Transport Cloud is compatible with other systems that I use (FC2) A specific person or group is available for assistance with difficulties concerning the use of the Transport Cloud (FC3) I found the data I was looking for (FC4)	Five-point Likert scale (strongly disagree–strongly agree)
Behavioural intention (BI)	I intend to use the Transport Cloud in the future (BI1) I predict that I will use the Transport Cloud in the future (BI2)	Five-point Likert scale (strongly disagree–strongly agree)
Voluntariness of use (VU)	Although it might be helpful, using the Transport Cloud is certainly not compulsory for my research or other activities (VU1) My research and other activities do not require me to use open public sector data (VU2) My use of Transport Cloud is voluntary (it is not required by my superior/research/other activities) (VU3)	Five-point Likert scale (strongly disagree–strongly agree)
Gender	Are you male or female? (G)	Multiple choice (male or female)
Age of respondent (AR)	What is your age? (AR)	Eight-point scale (under 18–61 or over)
Purpose of use	To what extent are the following purposes important for your use of mobility open public sector data? (P)	Five-point Likert scale (very unimportant–very important)

Funded by the European Union

Type of data (T)	Which of the following types of mobility data did you find or have you used? (T)	Multiple choice (type of public sector data: geographic, legal, meteorological, social, transport, business, other, namely …)
Condition of data (C)	In which condition were the mobility data?	Five-point Likert scale (Very bad– very good)
Age of data (AD)	Were mobility data too old?	Five-point Likert scale (Very old – very updated)

Furthermore, relevant to the transport Cloud is the Open Data Survey which is a tool that allows assessing the status of cities and other local areas. It is run since 2012 by Open Knowledge international and it produces the **Global Open Data Index (GODI)**. The objective is to assess open data publication and to understand how governments around the world publish open data and how much these data are useful for the public. The index also measures **findability.** Data assessment is based on four assumptions: 1. Data is defined according to the Open Definition which assumes that open data can be freely accessed, used, modified, and shared by anyone for any purpose. 2. Governments have the responsibility to publish data even if these data are managed by third parties. 3. GODI considers, not only government data publication but also the national publication of open data. It also considers the level of autonomy of local authorities from the government. If there is enough autonomy, the index is calculated on sub-national territories. 4. GODI ranks places, not countries.

The questions deployed to evaluate open data focus on **data usability** and **accessibility,** and they are reported in Table 13. Questions not scored are subjective assessment that depends on the respondent.

Table 13: Questions to calculate the Global Open Data Index

Question	Score
Is the data collected by the government (or a third-party related or linked to the government)?	Not scored
Is the data available online without the need to register or request access to the data?	15 points
Is the data available online at all?	Not scored
Is the data available free of charge?	15 points
Where did you find the data?	Not scored
How much do you agree with the following statement: “It was easy for me to find the data.”	Not scored

Is the data downloadable at once?	15 points
Data should be updated every [Time Interval]: Is the data up-to-date?	15 points
Is the data openly licensed/in the public domain?	20 points
Is the data in open and machine-readable file formats?	20 points
How much human effort is required to use the data? (1 = little to no effort is required, 3 = extensive effort is required)	Not scored

Finally, the survey for the evaluation of the acceptance of the Transport Cloud has been enriched with questions formulated based on the non-functional requirements of the MobiDataLab Virtual Labs. Table 14 reports on the left the non-functional requirements of the

Table 14: Non-functional requirements of the MobiDataLab Virtual Labs and related questions on the acceptance of the Transport Cloud

Non-functional requirements	Question
Usability. The platform shall be able to achieve the required functional requirements effectively and time-efficiently. User satisfaction shall be considered and feedback from Living Labs occasions shall be incorporated into the development.	Do you feel that the Transport Cloud is easy to use?
Accessibility. The platform shall focus on allowing users to easily access information and datasets resulting in promoting active engagement in data sharing. Extra attention shall be given to the integration between the Virtual Lab and Transport Cloud, since the information accessibility of the first is profoundly linked with the second.	Do you feel that it is easy to access the transport Cloud?
Digital inclusion. The platform shall promote access and use of digital mobility services and data by vulnerable and excluded users (i.e. 'non-digital natives', people with reduced mobility).	Do you feel digital inclusion is achieved through transport cloud/virtual lab?
Engage in co-creation. The platform shall aim to promote pair-working between users. The pair working shall be: 1) Between various Data Providers & Owners in the co-creation of a Virtual Living Lab, 2) Between Innovators/data users through cooperation during the problem-solving phase and joint submission.	Do you feel engaged in the co-creation of the Transport Cloud?
Adaptability. Since the implementation of the platform will incorporate feedback derived from the Living Labs occasions, the platform shall be easily modified and upgraded based on the given comments and improvement suggestions.	Do you feel that the Transport Cloud is easily adaptable to your improvement suggestions?

Funded by the European Union

Performance. The platform shall have a fast response time as it will be used during real-time activities (i.e. Virtual Living Labs). Response time shall be tracked throughout the project and carefully examined in order not to have extensive waiting periods.	Do you feel that the Transport Cloud have a fast response time?
Security. The platform shall ensure GDPR compliance throughout user's interaction (from filling personal information up to exchanging messages with other users). Special attention shall be given when submitting a solution regarding IPR and user rights. Additional attention will be given to the security of the databases which may include firewalls to prevent unauthorized access and possible malicious actions.	Do you feel the security of data is covered during the use of the Transport Cloud?
Reliability. The platform shall be able to operate regardless of the users' number and workload. Efficient data distribution shall be ensured throughout users' interaction. The user shall be informed when completing an asynchronous and heavy task in order to attain reliability as well as user satisfaction.	Are you satisfied with the use of the Transport Cloud?
Portability – Compatibility. The platform shall be visible and usable in any digital environment. Since Virtual Lab is promoting remote participation and engagement, users may utilize different devices.	Is the Transport Cloud compatible and usable with your device?
Scalability / Multi-contextuality Although the Virtual Lab will initially be used for promoting data sharing for mobility in Europe and tested/improved through the Living Lab instances, the ultimate goal of its implementation is the effective improvement in the culture of data sharing in general. The platform shall be built in a way that could easily be upgraded to include more areas of expertise that would benefit from data sharing (e.g. logistics).	Do you think that MobiDataLab improves the data sharing culture?

The survey to assess the acceptance of the Transport Cloud based on the UTAUT, the Global Open Data Index and the no- functional requirements of the Virtual Labs is reported in Table 15. Questions in bold have been integrated into the UTAUT questionnaire based on the Global Open Data Index and the non-functional requirements.

Table 15: Survey to assess the intention to use and accept the Transport Cloud

UTAUT construct	Question	Type of outcome
Performance expectancy (PE)	Using the Transport Cloud is of benefit to me (PE1)	Five-point Likert scale (strongly

	Using the Transport Cloud will enable me to accomplish my research more quickly (PE2) The possibility of easily finding mobility data thanks to the Transport Cloud will increase my productivity (PE3) The possibility of easily finding open data thanks to the transport Cloud will increase my performance in my job (PE4) Do you think that the Transport Cloud will have a fast response time? (**PE5**)	disagree–strongly agree)
Effort expectancy (EE)	It will be easy for me to become skilful at using the Transport Cloud (EE1) Learning to use the Transport Cloud will be easy for me (**EE2**) I clearly understand how to use the Transport Cloud (EE3) I do not have difficulties in explaining why using the Transport Cloud may be beneficial (EE4) To use the data is required a high human effort? (**EE5**)	Five-point Likert scale (strongly disagree–strongly agree)
Social influence (SI)	Do you feel digital inclusion is achieved through transport cloud/virtual lab? (**SI1**) Do you feel the security of data is covered during the use of Transport Cloud? (**SI2**) Do you think that the Transport Cloud improves the data sharing culture? (**S3**)	Five-point Likert scale (strongly disagree–strongly agree)
Facilitating conditions (FC)	I have the resources necessary to use the Transport Cloud (FC1) The Transport Cloud is compatible with other systems that I use (**FC2**) A specific person or group is available for assistance with difficulties concerning the use of the Transport Cloud (FC3) I found the data I was looking for (FC4) Do you feel that the Transport Cloud is easily adaptable to your improvement suggestions (**FC5**)?	Five-point Likert scale (strongly disagree–strongly agree)

Funded by the European Union

Behavioural intention (BI)	I intend to use the Transport Cloud in the future (BI1) I predict that I will use the Transport Cloud in the future (BI2) Are you satisfied with the concept of the Transport Cloud? (BI3)	Five-point Likert scale (strongly disagree–strongly agree)
Voluntariness of use (VU)	Although it might be helpful, using the Transport Cloud is certainly not compulsory for my research or other activities (VU1) My research and other activities do not require me to use open public sector data (VU2) My use of Transport Cloud is voluntary (it is not required by my superior/research/other activities) (VU3)	Five-point Likert scale (strongly disagree–strongly agree)
Gender	Are you male or female? (G)	Multiple choice (male or female)
Age of respondent (AR)	What is your age? (AR)	Eight-point scale (under 18–61 or over)
Type of data (T)	Which types of mobility data do you need? (T1) Is the data collected by the government (or a third-party related or linked to the government)? (**T2**) Is the data available online without the need to register or request access to the data? (**T3**) Is the data available free of charge? (**T4**) Is the data downloadable at once? (**T5**) Is the data in open and machine-readable file formats? (**T6**)	Multiple choice (type of public sector data: geographic, legal, meteorological, social, transport, business, other, namely …) Yes/No
Condition of data (C)	In which condition usually are the mobility data? (C1)	Five-point Likert scale (Very bad–very good)
Age of data (AD)	Are mobility data usually too old? (AD1) Is the data you need up-to-date? (**AD2**)	Five-point Likert scale (Very old – very updated)

The interviews to evaluate the acceptance of the Transport Cloud will be also performed with MobiDataLab events participants. In this case, it is interesting to have interviews with different types of respondents such as public

Funded by the European Union

7. Conclusions

In this deliverable, the methodology to assess the impact of the data sharing culture on the society and environment has been presented. Since the Transport Cloud and the Business Models are expected to have an impact on the mobility data sharing culture, the framework comprises the evaluation of the acceptance of the Transport Cloud and the Business Models.

Initially, the areas of impact of the data sharing culture have been analysed. These impact areas are the usage of transport data, the economy, the society, and the environment. Regarding the usage of transport data, several digital transport services have been identified and described in terms of their capability to produce an impact on the environment or society. Concerning the impact on the economy, the data sharing culture is expected to create the possibility to develop new innovative services but also to create new direct and indirect jobs and increase revenues. However, public organizations should be aware of the advantages of sharing their data and should understand which data is more convenient for them to share. In any case, there are still problems with interoperability and there is a need to ensure privacy. Several works claim that the data sharing culture can help reduce the impact of transport on the environment, and improve safety, quality of life and social conditions. However, the impact on the society and environment mainly depends on the type of application or service and data that are deployed. Surely, in the case of digital transport systems, the more data is integrated the most impactful is the device on the travellers' behaviour.

The use cases identified in D2.9 are discussed in terms of their capability to have a positive impact on society or reduce the impact of transport on the environment. Overall, there are several ways in which the use cases can have a socio-environmental impact. These ways will be explored thanks to questions formulated based on the hypothesis of impacts. The interviews will be held with the project participants and will provide an overview of how the Use Cases impact the environment and society.

A set of KPIs identified thanks to the analysis of similar existing initiatives, will be selected, and adapted to the MobiDataLab context. These KPIs will be not measured because the Transport Cloud will be only a prototype. However, these measures are important because they can assess the acceptance of the Transport Cloud by the different type of stakeholders.

The last part of the deliverable is dedicated to the evaluation of the acceptance of the Business Models and the Transport Cloud. The surveys will be executed during MobiDataLab events through interviews performed with participants in the context of T5.2. However, if the interviewed participants will not completely represent the possible stakeholders or users of the Transport Cloud or possible adopters of the Business Models, additional participants will be asked to answer the survey.

Funded by the European Union

Bibliography

Abraham, R., Schneider, J. and vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda, International Journal of Information Management, 49, 424-438, https://doi.org/10.1016/j.ijinfomgt.2019.07.008.

Atz, U. (2014). The Tau of Data: A New Metric to Assess the Timeliness of Data in Catalogues. In Proceedings of the International Conference for E-Democracy and Open Government (CeDEM2014), Krems, Austria.

BLUE CLOUD, www.blue-cloud.org

Bhati, B. S., Ivanchev, J., Bojic, I., Datta, A. and Eckhoff, D. (2021). Utility-Driven k-Anonymization of Public Transport User Data, IEEE Access, 9, 23608-23623. doi: 10.1109/ACCESS.2021.3055505.

Boston Consulting Group, *www.bcg.com*

Budsky, P. How to evaluate business model? Available at: https://stc.fs.cvut.cz/pdf19/9546.pdf

Caballero-Gil, C., Molina-Gil, J., Hernández-Serrano, J., León, O., Soriano-Ibañez, M. (2016). Providing k-anonymity and revocation in ubiquitous VANETs, Ad Hoc Networks, 36(2), 482-494. https://doi.org/10.1016/j.adhoc.2015.05.016.

Castaneda, J., Cardona, J. F., Martins, L. D. C., & Juan, A. A. (2021). Supervised Machine Learning Algorithms for Measuring and Promoting Sustainable Transportation and Green Logistics. Transportation Research Procedia, 58, 455-462.

Cheney, C. (2022). Tips for designing effective hackathons for social impact, 30/08/2017. https://www.devex.com/news/tips-for-designing-effective-hackathons-for-social-impact-90820 (Accessed on 5 april 2022).

Cheung, P., and Sengupta, U. (2016). Analysis of Journey Planner Apps and Best Practice Features.

Colpaert, P., Compernolle, M.V., Walravens, N., Mechant, P., Adriaenssens, J., Ongenae, F., Verborgh, R., and Mannens, E. (2017). Open transport data for maximising reuse in multimodal route planners: a study in Flanders. Iet Intelligent Transport Systems, 11, 397-402.

Datumize, www.datumize.com

Demir, E., Bektaş, T. and Laporte, G. (2011). A comparative analysis of several vehicle emission models for road freight transportation, Transportation Research Part D: Transport and Environment, 16, 5, 347-357, https://doi.org/10.1016/j.trd.2011.01.011.

Domingo-Ferrer, J. and Mateo-Sanz, J. M. (2002). Practical data-oriented microaggregation for statistical disclosure control," in IEEE Transactions on Knowledge and Data Engineering, 14(1), 189-201 doi: 10.1109/69.979982.

Domingo-Ferrer, J., Trujillo-Rasua, R. (2012). Microaggregation- and permutation-based anonymization of movement data, Information Sciences, 208, 55-80, https://doi.org/10.1016/j.ins.2012.04.015.

Ehret, M., Kashyap, V., & Wirtz, J. (2013). Business models: Impact on business markets and opportunities for marketing research. Industrial Marketing Management, 42(5), 649-655.

EOSC Secretary (2020). Open Consultation for the Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC). Retrieved from: https://www.eoscsecretariat.eu/sites/default/files/open_consultation_booklet_sria-eosc_20-july-2020.pdf. Last accessed 28 th February 2022.

Erlingsson, Ú., Korolova, A., & Pihur, V. (2014). RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security.

Fagerholt, K., Laporte, G. and Norstad, I. (2010). Reducing fuel emissions by optimizing speed on shipping routes, Journal of the Operational Research Society, 61:3, 523-529, DOI: 10.1057/jors.2009.77

Fattah, L., Gabrilove, J., and Bradley, F. (2021). Evaluating the impact of a health hackathon on collaborative team science: A Social Network Analysis (SNA). Journal Shaheenof Clinical and Translational Science, 5(1), E5. doi:10.1017/cts.2020.46

Giesecke, R., Surakka, T. and Hakonen, M., Conceptualising Mobility as a Service, 2016 Eleventh International Conference on Ecological Vehicles and Renewable Energies (EVER), 2016, 1-11, doi: 10.1109/EVER.2016.7476443.

Gilsing, R., Wilbik, A., Grefen, P. et al. (2021). Defining business model key performance indicators using intentional linguistic summaries. Softw Syst Model 20, 965–996. https://doi.org/10.1007/s10270-021-00894-x

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

Gordijn,J. and Akkermans, H. (2001). "Designing and evaluating e-business models," in IEEE Intelligent Systems, 16, 4, 11-17, doi: 10.1109/5254.941353.

Haaker, T., Bouwman, H., Janssen, W., & de Reuver, M. (2017). Business model stress testing: A practical approach to test the robustness of a business model. Futures, 89, 14-25.

Haaker, T., Bouwman, H., Janssen, W., de Reuver, M. (2017). Business model stress testing: A practical approach to test the robustness of a business model, Futures, 89, 14-25, https://doi.org/10.1016/j.futures.2017.04.003.

Hayes, J., & Finnegan, P. (2005). Assessing the of potential of e-business models: towards a framework for assisting decision-makers. European Journal of Operational Research, 160(2), 365-379.

Hidalgo, A. C., How data sharing is shaping the future of mobility, opendatasoft (04.02.2021). https://www.opendatasoft.com/en/blog/how-data sharing-is-shaping-the-future-of-mobility/ Accessed on 2 April 2022.

Joyce, A., and Paquin, R. L. (2016). The triple layered business model canvas: A tool to design more sustainable business models. Journal of cleaner production, 135, 1474-1486.

Koichi, S. (2020). Public Transport Promotion and Mobility-as-a-Service, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences 0916-8508, The Institute of Electronics, Information and Communication Engineers, 226-230 https://ci.nii.ac.jp/naid/130007778777/en/ 10.1587/transfun.2019TSI0001

Kujala, R., Weckström, C., Darst, R. et al. (2018). A collection of public transport network data sets for 25 cities. Sci Data 5, 180089. https://doi.org/10.1038/sdata.2018.89

Lind, M., Alvarado, J. L., haraldson, S., Mulder, H., Nykanen, L. and Piccoli, G. Digital data sharing for greener transport in sustainable supply chains – the benefits of establishing a Federated network of platforms (2/6/2021). https://trans.info/en/digital-data sharing-for-greener-transport-in-sustainable-supply-chains-the-benefits-of-establishing-a-federated-network-of-platforms-239539 Accessed on 3 April 2022.

Lizarralde, I., Hamwi, M., Abi Akle, A., Samir, B., Lentzen, L. and Kromrey, V. (2020). SOCIALRES: Comparative analysis of existing business models for RES cooperative, aggregators and crow-funders. Deliverable of the SOCIALRES European project (more information can be found here: https://socialres.eu/project/

Loder, A., L. Ambühl, M. Menendez and K.W. Axhausen (2019). Understanding traffic capacity of urban networks, Scientific Reports, 9 (1) 16283. https://doi.org/10.1038/s41598-019-51539-5

Lucas, K. (2012). Transport and social exclusion: Where are we now? Transport policy, 20, 105-113.

MaaS Alliance (2017) Guidelines & Recommendations to create the foundations for a thriving MaaS Ecosystem, White Paper.

Mansour, M., Fahmy, A. and Hashem, M. (2012). Maintaining location privacy and anonymity for vehicle's drivers in VANET, International Journal of Emerging Technology and Advanced Engineering 2(11) 8–40.

Meurisch , C. and Mühlhäuser, M. (2021). Data Protection in AI Services: A Survey. ACM Comput. Surv. 54, 2, Article 40 (March 2022), 38 pages. DOI:https://doi.org/10.1145/3440754

OECD https://www.oecd-ilibrary.org//sites/90ebc73d-en/index.html?itemId=/content/component/90ebc73d-en# Accessed on 8 April 2022.

OpenAIRE-Connect - CONNECTing scientific results in support of Open Science, www.openaire.eu

OpenDataMonitor, project.opendatamonitor.eu

Open Data Survey, census.okfn.org

Osterwalder, A. (2004). The business model ontology a proposition in a design science approach (Doctoral dissertation, Université de Lausanne, Faculté des hautes études commerciales).

Osterwalder, A. (2007). How to describe and improve your business model to compete better. draft version, 8.

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

Perboli, G. (2017). The GUEST methodology. Available at: https://staff.polito.it/guido.perboli/GUEST-site/docs/GUEST_Metodology_ENG.pdf

Poom , A., Helle, J. and Toivonen , T. (2020). Journey planners can promote active, healthy and sustainable urban travel. Urbaria Summaries Series, 6 ,2020, Helsingin yliopisto, Kaupunkitutkimusinstituutti Urbaria.

Samarati, P. and Sweeney, L. (2018). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Carnegie Mellon University. Journal contribution. https://doi.org/10.1184/R1/6625469.v1

Shaheen, S., and Cohen, A. (2020). Mobility on demand (MOD) and mobility as a service (MaaS): Early understanding of shared mobility impacts and public transit partnerships. In Demand for emerging transportation systems, 37-59. Elsevier.

Sochor, J., Arby, H., Karlsson, M. and Sarasini, S. (2017). A topological approach to Mobility as a Service: A proposed tool for understanding requirements and effects, and for aiding the integration of societal goals. 1st International Conference on Mobility as a Service (ICOMaaS), Tampere, Finland.

Stanley, J. K., Hensher, D. A., Stanley, J. R., & Vella-Brodrick, D. (2021). Valuing changes in wellbeing and its relevance for transport policy. Transport policy, 110, 16-27.

Studer, A., Shi, E., Bai, F. and Perrig, A. (2009). Tacking together efficient authentication, revocation, and privacy in VANETs, in: Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON '09. 6th Annual IEEE Communications Society Conference on, 1–9, doi: 10.1109/SAHCN.2009.5168976.

Szopinski, D., Schoormann, T., John, T., Knackstedt, R., and Kundisch, D. (2020). Software tools for business model innovation: current state and future challenges. Electronic Markets, 30(3), 469-494.

Transport Observatory / fOrum for Promoting Open Science, TOPOS, www.topos-observatory.eu

Tsohou, A., Magkos, M., Mouratidis, H., Chrysoloras, G., Piras, L., Pavlidis, M., Debussche, J., Rotoloni, M., Crespo Gallego-Nicasio, B., Katsikas, S., Cuppens, F., Cuppens, N., Lambrinoudakis, C., Kalloniatis, C., Mylopoulos, J., Antón, A., Gritzalis, S., Pallas, F., Pohle, J., Sasse, A., Meng, W., Furnell, S., Garcia-Alfaro, J. (2020). Privacy, Security, Legal and Technology Acceptance Requirements for a GDPR Compliance Platform, Computer Security, Springer International Publishing, 204-223.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540

Zuiderwijk, A., Janssen, M. and Dwivedi, Y.K. (2015). Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology, Government Information Quarterly, 32, 4, 429-440.

Zuurbier, M., Hoek, G., Oldenwening, M., Lenters, V., Meliefste, K., van den Hazel, P., Brunekreef, B. (2010). Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route. Environ. Health Perspect. 118, 783–789.

World Business Council for Sustainable Development (WBCSD), Enabling data sharing: Emerging principles for transforming urban mobility

Funded by the European Union

Annex I INTERVIEW PROTOCOL

COLLABORATION EFFECTIVENESS FOR QUALITY DATA SHARING, SOCIAL AND ENVIRONMENTAL IMPACT

INTRODUCTION:
Briefly explain purpose of the project and purpose of the interview.
State the duration of the interview
State the main parts of the interview (the main sections)
Ask for the permission to record the interview and explain that recording is used only for research purposes, to enable transcription of data, and ensure that no info is lost in the process of data gathering and analysis
State that data will be analyzed in an aggregate form, and that single responses and sources will remain anonymous.

Please introduce yourself and your involvement in the project, and in general your role in relation to mobility data sharing solutions.

SECTION A: COLLABORATION AND QUALITY OF KNOWLEDGE SHARING:
I would also like to ask you a couple of questions related to collaborations for data mobility integration in this project (1) and with your organizations' business partners and stakeholders in general (2). Depending on the stakeholder, 1 or 2 may be more relevant. Some questions apply only to 1, others to both.
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, and in such case, do you see them as complementary?
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 in this project have a shared vision of integrated mobility?

Would you say that you trust these stakeholders?
do you find them honest, reliable, and benevolent in attaining the project's goal (i.e., data sharing for integrated mobility)?
Do you think they will keep their promise related to this project (1) or to the goal of integrated data sharing for mobility (2)?
Would you say that you are satisfied with the quality of the data they share?
Would you say that you are satisfied with how they collaborate to provide the data?

- Are they contributing to developing standards and compliance mechanisms in the present or will they do so in the future?
- What is missing? What are the critical points?

Do you think that previous collaboration/knowledge sharing on mobility is a prerequisite for developing successful solutions for this project?
Partnership evolution processes: Would you say that there were different visions and expectations about the project in the beginning? Were there conflicts about envisioned solutions or work plans? Were these negotiated or solved as the project evolved?
Did you have social relations (i.e., do you know or have frequent business/personal contacts) with other project stakeholders before this project began?

- Do you think social relations impact the effectiveness of the data sharing in this project?
- What about social and environmental impact?

Do you think the partners are committed to this project? What about other important stakeholders that are not project partners, how committed do you think they are to sharing high quality information?

- Do you think that social relations among data providers and integrators can in any way influence the social and environmental impact of the envisioned solutions?

In relation to your use case, what kind of collaboration orchestration is necessary to ensure successful results?

- In your experience (and in particular reference to your use case(s)), would you say that there is a leader or orchestrator in the network of data exchanges? Do you think it is a necessary function? Do you think network/collaboration orchestration can help towards the measurement of social and environmental impact? Specifically, do you think that some partners which have a higher focus on societal and environmental impact should lead the way?

Do you feel that the project's stakeholders have managed well deadlines so far?

- In case there were issues with deadlines, which were this?
- Were there any issues with deadlines for measuring social impact and dissemination?
- Out of 100%, what % of time was spent coordinating about operational tasks and what % about dissemination and social impact goals?

<u>SECTION B: QUESTIONS TO SPECIFICALLY ASSESS DATA SHARING AND COLLABORATION CULTURE:</u>

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?

- ONLY IF NECESSARY (they stay very generic or elusive), ask questions about: interoperability to avoid fragmentation, openness of data, reusability, subsystem transferability, security and privacy.

How would you rate your current level of collaboration with your mobility business partners and stakeholders? From 1 to 10, how would you rank your business process integration? Also, how would you rank the business process integration of your partners?
How much do you currently work on continuously improving info sharing with your partners/? Are there any projects that you are involved in?
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?

Funded by the European Union

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?

SECTION C: SOCIAL IMPACT:
Do you think that this project (1) and integrated mobility solutions in general (2) can bring systemic social and environmental change?
Let's first talk about the social aspect: What is social change to you and what kind of social change can integrate 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?

Who are the main beneficiaries of this project and of similar solutions, according to you? How can their lives be improved?
What do you think are the main challenges of this project (or of integrated mobility solutions in general) for providing social impact?

- What do you think is more important, advanced technological solutions for data management and integration, or collaboration in extended data mobility networks? If both are important, how can they be interconnected to achieve impactful solutions for societies and the environment?

What do you think are the main challenges of this project (and of similar projects aiming at mobility data integration) in terms of measuring social and environmental impact and providing adequate KPIs? Are there any courses of action that you have adopted or plan to adopt in order to deal with these challenges?

SECTION D: SPECIFIC QUESTIONS ON THE PERCEIVED IMPACT OF USE CASES
ETA:
Actors: Main: Dispatcher, Transport customer, Fleet owner. Secondary: data sharing platform and fleet IT system. -will these actors be involved in the interviews?
How can optimization for freight deliveries with estimated time of arrival reduce the impact of transport on the environment?
How, in your opinion, the optimization of transport flows with ETA is able to create a positive impact on the quality of work of the freight delivery of company employees?
In which way, in your opinion, the route optimization for freight deliveries with estimated time of arrival is able to improve the social conditions of citizens?
How can optimization of transport flows with ETA inform policymakers?
In each of these, what is the most critical factor for an impactful ETA implementation? (only if necessary mention alerts for delayed stops, sharing arrival time, post trip reporting, update tour plan, rest time planning)

EMISSION REPORTING:
Actors: Main: Dispatcher, Transport customer, Fleet owner. Secondary: data sharing platform and fleet IT system. -will these actors be involved in the interviews?

How, in your opinion, the emission reporting helps to reduce the impact of transport on the environment?
How can the emission reporting produce a positive impact on the society?
We know that emission reporting is a complex business. What are the main critical aspects of emission reporting, according to you? (Reporting indirect and direct emissions, attributing emissions?)
How can we overcome the challenges of emission reporting to maximize social impact? What about environmental impact?
To which extent do you think stakeholders in this project have already good practices in terms of report emissions, report predicted emissions, and history of past emissions?
To which extent do you think there is willingness and commitment in the network to develop common standards and practices for report emissions?
To which extent do you think there is transparency and honesty in report emissions?

ANALYTICS AND MACHINE LEARNING:
Actors: Primary: Researcher, the Data Scientist and the Domain expert, Secondary: the Decision Maker will these actors be involved in the interviews?

If the questions in the previous sections have been addressed, there may be no need for this part on data quality: 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 designing a reliable journey planner?

- ➢ 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?
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?

REUSE OF TRANSPORT DATA
Actors: Primary Digital Service Provider, the Public Transport Authority, the Public Transport Operator, the contributors and the mobility operators; Secondary: the External calculator
Raw transport datasets are often too complicated to allow different types of stakeholders to offer journey planning services. A common solution for journey planning would foster the use of transport data, through a service layer (unified API, common user experience) that simplifies the usage and allows the use also for non-domain experts.
How, in your opinion, could the journey planner create positive social and environmental impact?

Funded by the European Union

- Only if they are vague ask about preventing the exposure of travelers to environmental pollutions, reducing the impact on the environment of each single trip? More attention to the exposure of cyclist and pedestrians?

MOBILITY AS A SERVICE (MAAS)
The MaaS is an "integration of various forms of transport services into single mobility service accessible on-demand" (based on the journey planner): retrieve transport data, plan journeys.
Actors: MobiDataLab user
How does MaaS contribute to foster multimodal mobility and the use of new forms of mobility that allow to reduce the impact of transport on the environment? (less congestion, pollution?)
How can MaaS improve the liveability of cities?
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?
What could be some ways to increase social attention and relevance of the project at the light of these social shocks?

NB: bullet points questions are intended as support for the main questions, to be used according to situational context of each interview

Annex II : Data Protection

In this Annex, a review of data protection techniques is presented. This can be considered as an integration of the review of data protection techniques presented in D2.3. Furthermore, this chapter is helpful to provide an overview of the existing techniques to protect privacy thus overcoming one of the barriers of data sharing, as already mentioned in Chapter 2 describing the economy impact area.

Privacy which can be ensured thanks to the implementation of data protection techniques. Data protection techniques have been developed to anonymise different types of data. For what concerns trajectory data, traditional techniques for data anonymization cannot be applied because of computational complexity. In this regard, Domingo-Ferrer and Trujillo-Rasua (2012) propose two heuristics for trajectory anonymization. The authors claim that the proposed technique allows to reach k-anonymity. To ensure the anonymity of individuals, **k-anonymity** technique allows to map each observation ambiguously to at least k entities. Samarati and Sweeney (2018) introduce the **generalization** and **suppression** techniques. Generalization consists of changing the domain of a set of values. The domain in relational database is defined as a set of values that an attribute can assume. For instance, if the age of a person in 46, the information could be generalized into the class of ages included between 40 and 49. The generalization relationship is the mapping between the true value and the new generalized domain. Suppression is a complementary approach to generalization which consists of removing data. It is applied when generalization is not possible. Furthermore, **substitution** (**or perturbation**) consists of replacing original data with perturbed information.

Caballero-Gil et al. (2016) consider Vehicle Ad hoc Networks (VANET) which combines data from the cloud. VANET consists of On-Board Units (OBU) combined with Road Side Units (RSU) which are placed at fix locations on the road infrastructure. Traffic management is carried out by considering the location of the vehicles but, privacy cannot be ensured. The authors propose an approach that guarantees k-anonymity but it is able to identify malicious users. There are several other works that tackle the privacy problem of VANETs by, for instance, confusing trajectories (Mansour et al. 2012) or by providing temporary anonymous keys (Studer et al. 2009). However, this second solutions only partially solves the privacy problem because the traffic managers know the real identities.

Domingo-Ferrer and Mateo-Sanz (2002) proposed the **microaggregation** technique to reach k-anonymity. The method consists of aggregating individual records (microdata) into small groups which includes at least k records. The idea is to replace individual values with the values computed on the aggregate. The groups of data are formed based on their similarity. The microaggregation approach is therefore strictly related to the clustering problem.

Meurisch and Mühlhäuser (2021) review the main area in which data protection is very important in Artificial Intelligence (AI) services. The paper is very interesting especially in the light of UC3 Analytics and Learning. The authors classify data protection techniques based on different points of views:

1. Protection at manager level. The point of view of the AI service manager who needs to protect data from unauthorised access when stored. Here the problem is also related to the output of the AI service that is completely based on user data and stored elsewhere without user's control. In this case, to ensure user privacy, it is necessary to *authenticate* and *authorize* the

user or the AI service. The system should also be able to ensure the *consistency* of data and the confidentiality. There are several protections approach that are used in centralised database to control access which are implemented by security experts. Decentralised approaches to access control usually require the user to define the rules manually. More user-friendly approaches identified by the authors are: KP-ABE which allows to share encrypted data; decentralised approaches based on blockchain (ENIGMA, ControlChain), privacy preserving mechanisms: SWYSWYK (Share What You See With Who You Know), Data Cooperatives in which individuals pool their data for the benefit of the membership of the group or community and Personal Data Stores (PDS) such as Google Drive or Microsoft OneDrive consisting of several implementations such as Personal Data Vault (PDV) which only supports location data, P3 which only supports JPEG, OpenPDS allows to aggregate user data but it does not guarantee anonymity. The authors list several other solutions which, however, have not been developed either at prototype and point out that the main difficulty is to ensure data storage and privacy and independence from a platform.

2. Protection at system level. In this case, users deploy data in the cloud or AI services hosted by third parties. Therefore, privacy should also be ensured on this cloud or third party's storages. The authors categorise data protection approaches in two subcategories of software based and hardware-based virtualisation. Here the main issue is to ensure data protection while ensuring low latency for real time data transmission.

3. Protection at AI level. The third case is related to data transfer which is needed because AI services access data to provide personalised services. Protection approaches are here classified into several subcategories.

    - Data modifying approaches. These are **k-anonymity** and the variants (l-diversity, t-closeness); **perturbation/obfuscation** (differential privacy such as RAPPOR from Google, SuLQ, Incognito, Obfuscation At-source, and Pickle);

    - Data-encrypting approaches. There are two complementary data encryption techniques. **Homomorphic encryption** (HE) that allows to analyse and manipulate data but it is inefficient form the computational point of view and **Secure multi-party computation** (MPC) which is a cryptographic protocol that allows to perform computation on distributed data.

    - Data minimizing approaches. It consists of creating **general model (GM)** for training which usually do not require personal information. Another example is the **collaborative learning** of a shared model such as the Federated Learning (FL) from Google or When Edge Meets Learning. Here the user learns a personal model, the provider learns and adapt this general model by aggregating models.

Data encryption standard (DES) is a symmetric algorithm that has been developed by IBM in the 70s. Symmetric means that from the encrypted data it is possible to return to the original data using a keyword. Similarly, advanced encryption standard (AES) encryption, or is the encryption methodology adopted by the US government to protect documents. The difference between the two approaches relies on the block size, that is the number of bits of the fixed length string.

With reference to **Homomorphic encryption** (HE), this is a standard which allows to perform the computation directly on data, there are several open-source implementations. Some of them are

Funded by the European Union

listed in the website homomorphicencryption.org. Here below is a selection of open sources projects that implement the HE:

- Microsoft SEAL (github.com/Microsoft/SEAL) is written in C++ and it provides a simple API together with several example for its usage. Compared to the other approaches, thanks to Microsoft SEAL the cloud operator will never have unencrypted access to customer data.
- PALISADE (palisade-crypto.org) is supported by the US Defense Advanced Research Projects Agency (DARPA). It supports various homomorphic schemes (BGV, BFV, CKKS, TFHE and FHEW).
- HELib (github.com/shaih/Helib) supports Brakerski-Gentry-Vaikuntanathan (BGV) scheme with bootstrapping and the Approximate Number scheme of Cheon-Kim-Kim-Song (CKKS)
- FHEW/TFHE is based on the Fully Homorphic Encryption scheme and it makes use of the library Fastest Fourier Transform in the West (FFTW).

Erlingsson et al. (2014) proposed the Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR) which is based on the *randomized response*, a technique of the 60s which consists of asking the respondent to flip a coin and to decide to answer the truth to the sensitive question or to answer the question based on the flipped coin results.

Bhati et al. (2021) focus on transport data and propose an approach to obtain k-anonymity thanks to normalization techniques, distance and utility metrics which can be deployed with numerical or categorical attributes. The final objective is to minimize the distortion of the data. The authors test the methodology on a real dataset and show that it is sometime better to suppress a record to preserve the utility.

In the context of OpenAIRE, AMNESIA (amnesia.openaire.eu/amnesia) is an online tool that allows to automatically anonymise dataset. SharedStreets creates open-source software for aggregation and encryption that protects user privacy and business interests when dealing with potentially identifiable information.

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).