MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

FINAL BROCHURE

February 2024

https://mobidatalab.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101006879. The content of this article reflects solely the views of its authors.

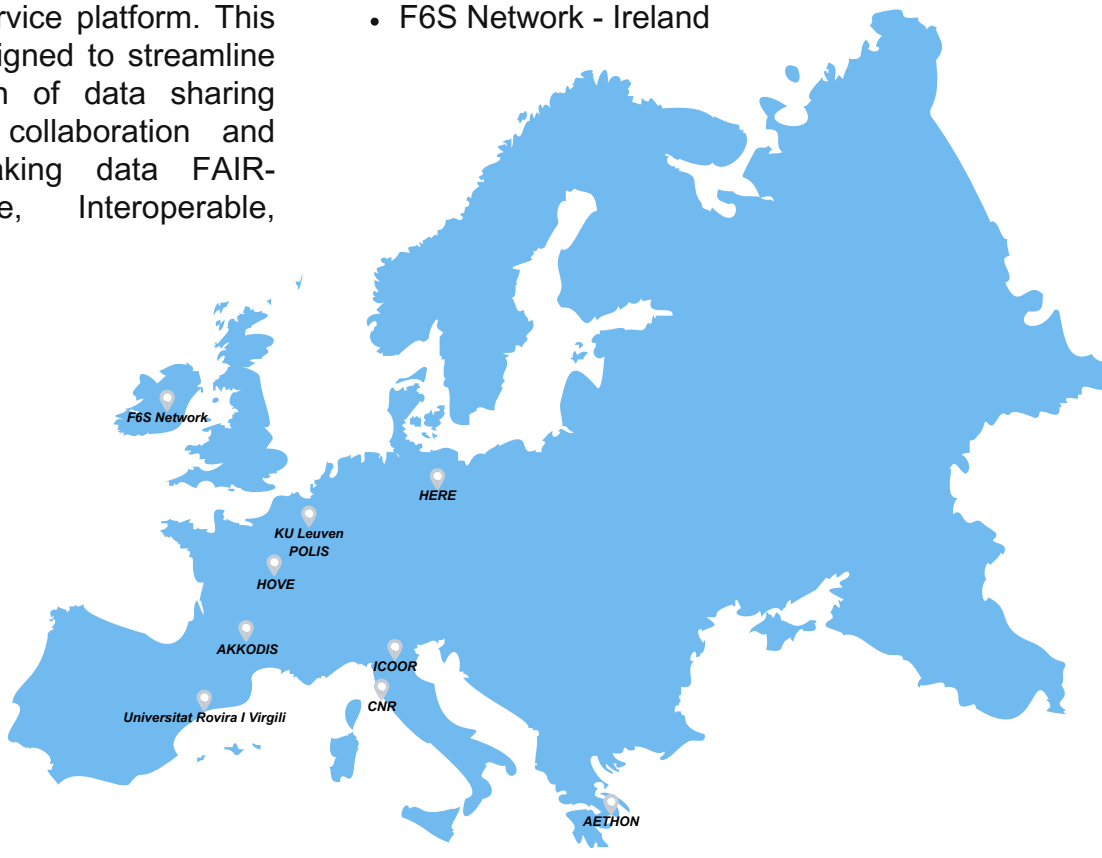
1. MOBIDATALAB IN A NUTSHELL

MobiDataLab was an EU-funded project in the framework of the Horizon 2020 research programme, running from January 2021 to January 2024, and dedicated to fostering a culture of data sharing among mobility stakeholders. The objective of the project was to introduce stakeholders (including transport organizing authorities, operators, industry, and innovators) to a methodology and tools aimed at promoting the advancement of data sharing practices in Europe and beyond.

At the heart of MobiDataLab lied a commitment to continuous co-development of knowledge and technical solutions, engaging all stakeholders in the transport and mobility landscape. We facilitated problem-solving Labs, collected and analysed expert advice and recommendations, and supported cities, regions, industry, and associations by building a cross-thematic knowledge base and a cloud-based service platform. This infrastructure was designed to streamline access and utilization of data sharing resources, fostering collaboration and innovation while making data FAIR-Findable, Accessible, Interoperable, Reusable.

MobiDataLab brought together a consortium of 10 partners from various sectors, including Industry, Research, Academia, Consultancy, and Governance, spanning 7 countries. This diverse consortium shares a common commitment to open data and open-source principles, with a focus on scalability and replicability beyond the EU.

- AKKODIS - France
- Consorzio Interuniversitario Per L'ottimizzazione E La Ricerca Operativa (ICOOR) - Italy
- AETHON - Greece
- Consiglio Nazionale Delle Ricerche (CNR) - Italy
- Hove - France
- HERE - Germany, Netherlands
- Katholieke Universiteit Leuven/KU Leuven (KUL) - Belgium
- Universitat Rovira I Virgili (URV) - Spain
- POLIS - Belgium
- F6S Network - Ireland



2. THE MOBIDATALAB APPROACH

The MobiDataLab approach rested upon four foundational pillars: the Open Knowledge Base, the Transport Cloud, the Labs (Living and Virtual) and the impact assessment on the data sharing culture.

1. The **MobiDataLab Open Knowledge Base** provides a user-friendly web-based tool that served as a repository of resources related to data sharing challenges and best practices. This platform curated insights from leading projects in the field, offering mobility data providers tailored recommendations on improving data quality, accessibility, and usability across various sharing models and environments. Enhanced with real-world use cases and mobility stakeholder input, the Open Knowledge Base became a go-to reference for navigating interoperability, legal frameworks, privacy concerns, and technological hurdles.
2. The **Transport Cloud** demonstrates a cloud-based prototype platform designed to facilitate seamless sharing of transport data among stakeholders. Evaluated by a Reference Group comprising data providers and users, this platform embraced federated cloud principles to offer a unified interface for accessing internal and external resources. By demonstrating open, interoperable, and privacy-preserving data exchange, the Transport Cloud aimed to break down technical barriers hindering data reuse, showcasing solutions to enhance accessibility and innovation.
3. **Living and Virtual Labs** provided open innovation ecosystems for effective improvement in the culture of data sharing. Living Labs brought together local stakeholders for real-time data exchanges and co-creation sessions, addressing immediate challenges and specificities within territorial contexts. Meanwhile, Virtual Labs provided a digital gateway for remote participants, fostering engagement and collaboration in a virtual space. Together, these labs facilitated agile problem-solving and long-term innovation, leveraging the Transport Cloud to connect stakeholders with data innovators and solution providers, ultimately promoting the value of data reuse and exchange.
4. The **Impact Assessment of the Data Sharing Culture** evaluates the consequences of digitalization and data sharing in the transport sector. It assesses how these innovations can drive new business models, enhance societal benefits, and mitigate environmental impacts. Through rigorous assessment, MobiDataLab aimed to determine the feasibility of leveraging data sharing to address evolving user needs while fostering sustainable transport services and economic opportunities.

Through these pillars, MobiDataLab succeeded in fostering a dynamic ecosystem where stakeholders could collaborate, innovate, and advance the culture of data sharing in the mobility sector. The following sections will delve into each pillar of MobiDataLab's approach, providing detailed insights and references for further exploration.

3. MAIN OUTCOMES

3.1. Open Knowledge Base

3.1.1. Legal and Governance

Data are increasingly viewed as a commodity to be traded in their own rights. This has created a growing interest from policymakers in the creation of so-called “data markets” and “data spaces”, fostering data sharing.

However, problems arise on a legal and governance level. The European Commission is stepping away from an “ownership right” or general exclusive rights on data, moving towards data governance notions – which are still in their infancy - and introducing specific data legislation to facilitate data sharing and data access. At the same time, a variety of “legacy” legal frameworks, such as intellectual property rights and competition law, may still also apply to data and data transactions. This results in a complex and fragmented legal framework.

The University KU Leuven produced two reports analysing the current EU legal and regulatory frameworks for data sharing in the transport sector, identifying specific legal and regulatory gaps and providing relevant recommendations (where possible to address). Specific focus was given to the use case of Mobility-as-a-Service. Another document produced examined what is data governance, the different data governance mechanisms that could apply to MobiDataLab’s use cases, as well as the impact of the EU's new data laws, and provided relevant recommendations.

3.1.2. Mobility Data Privacy

Mobility data in its simplest form are data about individuals that include locations at specific times. Historical location data, in the form of data sets where each of the records corresponds to an individual and contains location data for some periods of time, are referred to as trajectory microdata sets. Such trajectory microdata sets are often of interest to transport authorities, operators, and other stakeholders to evaluate and improve their services, the state of the traffic, etc. and are therefore often made publicly available or shared.

Mobility data have a high degree of uniqueness and regularity. Uniqueness means that data from different individuals can be easily distinguished, especially at some specific locations. Trajectory regularity means that, for single individuals, their data follow periodic patterns. Because of these two defining characteristics, trajectory microdata is prone to privacy attacks on individual users and is difficult to anonymize.

The CRISES research group at URV has provided a theoretical and practical oriented overview of available privacy-by-design strategies, privacy risk assessment tools and privacy-preserving technologies that allow for the development of responsible, ethically aligned and GDPR-compliant tools for storing, sharing and analysis of mobility and transport data, subject to GDPR principles, such as purpose limitation.

3.1.3. Mobility Data Standards

Standardisation is essential when producing and exchanging (meta)data. Always-evolving standards can be evaluated by how findable, accessible, interoperable and reusable they are. MobiDataLab, produced two deliverables in this regard, a State of the art on Mobility Data sharing standards [1] and a Report on new Mobility Data sharing standards [2]. Through these, MobiDataLab identified the context, application and evolution of diverse standards; provided implementation support information; offered examples about how these standards were used by MobiDataLab and provided recommendations about how the mobility community can use them.

During the different stages of the project, MobiDataLab implemented mobility data (from public transport, micro-mobility, shared mobility, road traffic, journey planning, mobility as a service, etc.) and cross-domain standards related to new mobility data sharing. These standards served and guided the work of several tasks, including the Data Management Plan, the Open Knowledge Base, and more particularly the tasks related to the Transport Cloud development.

For instance, mobility data sharing standards were taken into consideration for the support of the Reference Data catalogue, while the horizontal standards for the exchange of data in the cloud were considered for Data Privacy and the Architecture Design of the cloud.

The best practices for sharing data on the web contributed to the work of the Data Access Services and Data Channels tasks and geodata sharing and semantic interoperability standards were considered for the Data Processors.

3.1.4. Actors' Needs

The Actors' Needs and Cooperation Framework report provided the foundations for the MobiDataLab technological developments through the identification of the following key aspects:

- Categorisation and ranking of the main actors that partake in mobility data sharing
- Most important types of data needed for exchange
- Use cases that foster data exchange
- Key issues in data exchange

A Delphi Survey distributed to more than 98 experts and retrieving valuable input from more than 50 responders in conjunction with an Innovators Survey collecting more than 130 responses, concluded in the following key findings.

The most important actors identified for data exchange are, in order of significance, public administrators, public transport operators, mobility service providers, and government transport agencies. The most critical types of data identified for exchange include:

- Real-time demand in public and private transport
- Demand analysis through mobility habits
- Geolocation data on parking availability
- Real-time schedule updates in public transport
- Delays and disruptions in public transport
- Traffic Data

[1] https://mobidatalab.eu/wp-content/uploads/2023/01/MobiDataLab-D2.4-StateOfTheArtOnMobilityDataSharingStandards-v2.1.pdf
[2] https://mobidatalab.eu/wp-content/uploads/2023/12/MobiDataLab-D2.5-ReportOnNewMobilityDataSharingStandards_v1.0.pdf

The most relevant use cases found to foster data exchange are:

REAL-TIME DATA SHARING ACROSS MODES FOR BETTER OPERATIONS

DECISION SUPPORT THROUGH DATA SHARING

BETTER JOURNEY PLANNING THROUGH THIRD-PARTY DATA INTEGRATION.

Finally, the key issues related to data exchange involve interoperability, reliability, security, policy and lack of cooperation culture.

3.1.5. Market Study & Gap Analysis

A study was developed to present a description and analysis of the main products, services, and platforms used in transport data sharing. Specifically, 37 products, services, and platforms were positioned according to the data provided or aggregated, and the components of the value chain offered.

The study has identified the 9 data types with higher potential of impact on the creation of innovative digital services (vehicle location, environment, maps, payment, vehicle usage, static infrastructure, dynamic infrastructure, ticketing, user-generated), as well as the 4 main components of the mobility data sharing value chain (generation, collection, analysis, and exchange).

Relevant conclusions and insights to consider are the following:

- Data generators (e.g., telecommunication companies, connected device manufacturers, mobility service providers) use various simultaneous distribution channels to share their data.
- Detailed information on data sources and price is mainly confidential.
- Data generators compete with each other to be a proxy for mobility.
- Aggregators are becoming distributors of mobility services.
- Some data generators aggregate other data to limit the competition from aggregators within both the mobility data-sharing market and the mobility market

3.1.6. Existing Cloud Solutions and Frameworks

The project identified open frameworks and solutions that could provide effective and efficient ideas, solutions, approaches and services to MobiDataLab stakeholders and facilitate their collaboration, including secure methods to selectively share and access the data subject to any specific combination of privacy/visibility rules and conditions.

Three dimensions were identified: (1) the solutions for data sharing, the European

This document aims to better understand how mobility data sharing happens from a business perspective and has the ambition to support the decision-making processes of data-owners, data users, and data-aggregators regarding opportunities to trade, share or use mobility data.

projects and initiatives, the mobility data platforms and technological federated cloud solutions; (2) the SWOT analysis to discuss the strength, weaknesses, threads and opportunities; finally, (3) the solutions for the Transport Cloud data processors.

3.1.7. Mobility Data Sharing Use Cases

The Mobility Data Sharing use cases were essential foundations within the project. These use cases underwent continuous updates, with two important milestones represented in the two versions of the use cases deliverables. The initial set of use-cases was designed and derived by the project group based on their business expertise and experience. This baseline helped to guide further outputs, like the transport cloud architecture, its components, the standards discussion, business and revenue models, along their requirements. The second version included the feedback from the reference group of cities, resulted in refined and extended use cases. This version was also strongly influenced by the first living labs setups and were the base for the city challenges presented to innovators during the living lab events.

These use cases are not static; they adapt to the agility of the mobility ecosystem. With the contributions of the project partners, the reference group of cities and the Living and Virtual Labs participants, the use cases evolved and were adjusted according to new requirements and recent technology. These stakeholders provided guidance for additional use cases and demonstrated the possibilities of the MobiDataLab ecosystem, the transport cloud with the data catalogues and its services. They enabled showcases and helped connect the diverse group of stakeholders driven by mobility data sharing.

3.1.8. Business and Revenue Models

A study was developed to have a global overview of the possible business models for the Future Transport Cloud (FTC). It highlights that the FTC will provide Findable, Accessible, Interoperable and Re-usable (FAIR) data. This data will be used by the data consumers to support short term and real time decisions (e.g. related to the management of transport modes) and medium-term decisions (such as urban planning). The FTC will also aggregate, combine, consolidate, and access data coming from different platforms (public and private transport companies, ticketing services, mapping services, weather companies, …) by ensuring that each of them can maintain or improve its business model.

After an analysis of the literature and a benchmark, 5 business model archetypes are identified:

- Open data and open source
- Two-sided (advertising or sponsorship)
- Market place
- Software as a service
- Barter

Through interviews and workshops, these models were then analysed in terms of feasibility, attractiveness, advantages, drawbacks and pre-requisites and a comparison is established.

After a deep analysis, 4 insights can be highlighted:

1. *Business models of the Future Transport Cloud are **dynamic** and **agile**. The business model for the beginning will not be necessarily the same than the business model for scaling the Future Transport Cloud.*

2. *The business model of the Future Transport Cloud is likely to be a mix between **several archetypal business models**. In particular, several revenue models may coexist according to the granularity and origin of data (e.g. different revenue models for real time and historical data, raw or aggregated data). Similarly, different models may be used for accessing the data and for using the solution.*

3. *According to **local specificities**, different business models may be implemented. Because of different local habits and local business organisations, as well as maturity level, some cities or region may use different business models for the Future Transport Cloud.*

4. *Business models are **perceived differently** according to stakeholders.*

3.2. The Transport Cloud Prototype

3.2.1. Metadata and services catalogues

MobiDataLab made available two metadata catalogues, CKAN and a GeoNetwork, which gathered more than 115k datasets from multiple data portals in a single place. These datasets cover transport and mobility categories as well as other categories that enrich the mobility datasets in the context of the local challenges of the project stakeholders. Data was made discoverable, findable and accessible through these catalogues not only by providing the data sources of the datasets harvested but also by providing information about the different organizations offering open, premium and privileged data. Both catalogues offer human-readable and machine-readable metadata by being **accessible via an API** and a **platform**. Both are interoperable and allow the reuse of datasets by mobility digital services like journey planners.

The **CKAN** catalogue complies with mobility data sharing standards such as **DCAT(JSON/RDF)**, **DCAT-AP**, **CSW**, **CSW INSPIRE** and **CKAN API**. Metadata is translated during the harvesting stage which makes it possible to display the metadata in a common language. Furthermore, the language of the interface content and the headings of the metadata fields can be changed.

The **GeoNetwork** catalogue, a GIS-specific solution, allows consistent and standardized data representation. It supports the following harvesting protocols: OGC CSW 2.0.2 and WFS, GeoNetwork, Geoportal, and ArcSDE.

*A **QGIS guide** [3] and a **demonstration** were made available as a tool to access mobility data and learn about available basic tools on QGIS. It provides information about how to access the MobiDataLab CKAN and GeoNetwork metadata catalogues.*

*Four documents were created to provide tools to help access and manage the different datasets available on CKAN with **Jupyter Notebook** [4]. These contain information about data standards (how to access and manipulate them) and how to visualize data with OSMnx and Matplotlib (including Folium) Python libraries.*

The MobiDataLab Service Catalog [5] promotes the discoverability and interoperability of mobility services by leveraging the OpenAPI Specification (OAS), enabling seamless integration to a wide range of tools and workflows, accelerating the development of mobility service applications. It comprises three open-source components: a catalogue of mobility service API documents [6], a website that showcases service APIs with query capabilities [7], and an automated process for contributing new API documents [8].

[3] https://github.com/MobiDataLab/QGIS-guide-CKAN-GeoNetwork-access
[4] https://github.com/MobiDataLab/CKAN-access-with-Jupyter-Notebook
[5] https://mobidatalab.github.io/mdl-catalog-ui/
[6] https://github.com/mobidatalab/mdl-catalog-api
[7] https://github.com/mobidatalab/mdl-catalog-ui
[8] https://github.com/MobiDataLab/mdl-catalog-deployment

To find more about how to discover, manage, visualize the data on these catalogues you can consult the following deliverables and repositories: **Reference Data Catalogue** [9][10] and **Data Access Services** [11][12].

[9] https://mobidatalab.eu/wp-content/uploads/2023/12/MobiDataLab-D4.4-ReferenceDataCatalogueV2_v1.0.pdf
[10] https://github.com/MobiDataLab/reference-data-catalogue-demo
[11] https://mobidatalab.eu/wp-content/uploads/2023/12/MobiDataLab-D4.6-DataAccessServicesV2_v1.0.pdf
[12] https://github.com/mobidatalab/mdl-catalog-ui

3.2.2. Journey Planning APIs

Common sense regarding mobility, from the general public's perception, would first designate journey planning tools as the most prized asset. The MobiDataLab project did include, among other tools, Navitia, to offer advanced journey planning functionalities to living lab participants, thanks to converging functionalities embedded in the same product:

Navitia *is designed with a fully configurable interface to interrogate the journey planner content by manually manipulating the standard options shown by default on the graphical user interface (GUI) and, if necessary, switching to advanced options where higher level of customisation is offered to the Navitia user. This user approach primarily caters to participants who do not need to demonstrate technical knowledge but can utilize the GUI effectively.*

That same level of interrogation was equally available as APIs, since the GUI is actually connected to the journey planner engine by APIs that are documented and included on the platform to the innovators discretion to include in application that will require journey planner functionality without the application editor to re develop its own Navitia equivalent.

During the x-athons sessions, two Navitia platforms were made available to the audience:

- **The generic platform** hosting the transit data for all the cities or territories taking part of the challenges submitted to the x-athons attendees.
- **A customisable Navitia** where one was free to drop its own GTFS/PBF content and challenge the validity of that customised set of data against the goal to reach as a response to the challenge resolved by the participant.

This journey planner deployment helped to identify improvement in city/territorial mobility case where a localised event would imply to re-route the traffic flow or the consequences of closing part of the transit network or traffic in a designated location.

3.2.3. Data Enrichment Processors

The CNR's semantic enrichment processor introduces advanced techniques for enriching mobility data with semantic dimensions. The processor addresses the need for context-rich mobility data, as this can lead to valuable new insights.

Its three-step pipeline combines various data sources to produce semantically enriched mobility data, and in formats (i.e., RDF knowledge graphs) that facilitate integration and analysis via SPARQL queries. Accessible through a user interface and a webAPI server, the processor has been designed with modularity, configurability, and extensibility in mind, hence allowing future customizations for diverse applications.

Another approach to enriching mobility data is geographic enrichment, which allows different datasets and APIs to be combined on the basis of a common location. The processor, available as open source, supports different mobility and GIS standards such as OSM, GTFS, GeoJSON, etc. Incorporating geographic information into mobility datasets provides valuable insights into spatial patterns, route preferences, and urban dynamics. This enrichment not only improves the accuracy of location-based services, but also advances urban planning, transport management, and environmental sustainability. Being able to combine mobility data and geographical context helps decision-makers to formulate informed policies, design efficient transport systems and respond effectively to society's changing needs.

The final version of the anonymization module includes 6 anonymization methods, 1 privacy-preserving analysis method and 5 methods for computing various utility and privacy metrics. It also provides a command line interface (CLI) that allows users to use all the module functionalities in a straightforward way. The module is also ready to be deployed in a server and to process requests via an API.

The anonymization module has been designed with a focus on modularity, where pseudonymization or anonymization methods can be built using different components dedicated to preprocessing, clustering, distance computation, aggregation, etc. The focus has been on facilitating the addition of new methods and components, with the aim of encouraging contributions from other researchers.

The module is available at https://github.com/MobiDataLab/mdl-anonymizer along with a detailed documentation.

3.2.4. Privacy and Anonymisation Mechanisms

Within the Mobidatalab project, the CRISES research group at URV developed an anonymization module that mainly allows users to anonymize a mobility dataset in a simple way. The module also allows users to perform some mobility analysis in a private-way, and to compute some utility and privacy measures on both the original and the anonymized datasets.

3.3. The Living and Virtual Labs

3.3.1. The Datathon

The MobiDataLab Datathon was the first event of the MobiDataLab Living and Virtual Labs series. The focus of the event was the analysis of the available mobility data to identify further requirements and gaps beyond the ones identified via the use cases' development and the communication with the reference group.

A subset of the reference group members prepared challenges that were proposed to the Living Lab attendees to solve. These challenges were related to the MobiDataLab use cases, where the reference group members saw the opportunity to solve them by sharing mobility data.

The Datathon was organized in three phases: planning, execution and review. Within the planning phase, the organization of the event happened. Online webinars and demos were organized and the project was presented at events to increase visibility and awareness. Within this phase, the cities of the reference group prepared their challenges and presented them during the webinars to the potential attendees. The execution of the main event took place in Berlin as a 1 ½ days event on site and online for remote participation. The submissions were evaluated by an expert committee, resulting in three winning solutions. The "Global Fusion Team" as winner of the Datathon was featured in the HERE blog.

The feedback for the event was very positive and set the ground for the following Living and Virtual Labs.

3.3.2. The Hackathon

As a continuation of the MobiDataLab project's initiatives to promote progress, forge new connections, foster innovative thought, and shed light on the profound implications of mobility data sharing, the second installment of the Living Labs, the Hackathon, was held in the pursuit of creative solutions for real-world mobility challenges.

The Hackathon served as an open invitation to innovators within the mobility domain, offering them the opportunity to come up with novel data analytics tools utilizing resources available within the MobiDataLab ecosystem, including the Transport Cloud, and guided by the challenges, as well as the input from the project's reference group and the consortium members.

Similar to the first living lab, the Hackathon was organized into three distinct phases: planning, execution, and review. During the planning phase, challenges were defined and developed collaboratively with the reference group, and more specifically with the mobility authorities of four different cities: Paris, Milan, Eindhoven, and Leuven. Evaluation criteria were established to assess the solutions. Prior to the event, a webinar was conducted to inform potential participants about the MobiDataLab project, the Hackathon, and the specific challenges they would tackle. As for the execution phase, it involved communication and promotion efforts to attract interested participants, with the event adopting a hybrid format that allowed both physical and online

participation. The event took place in Paris, serving as a hub for on-site activities.

Navigating the intricacies of organizing such an event and combining online and on-site participation, proved to be both a challenge and an opportunity. While this approach enabled broader geographic inclusivity and encouraged diverse participation, it also required careful coordination to ensure clarity in communication and expectations. It also served as a bridge between the consortium members, the participants and the field experts to network and to share their own ideas and feedback on the project prospectives.

3.3.3. The Codagon

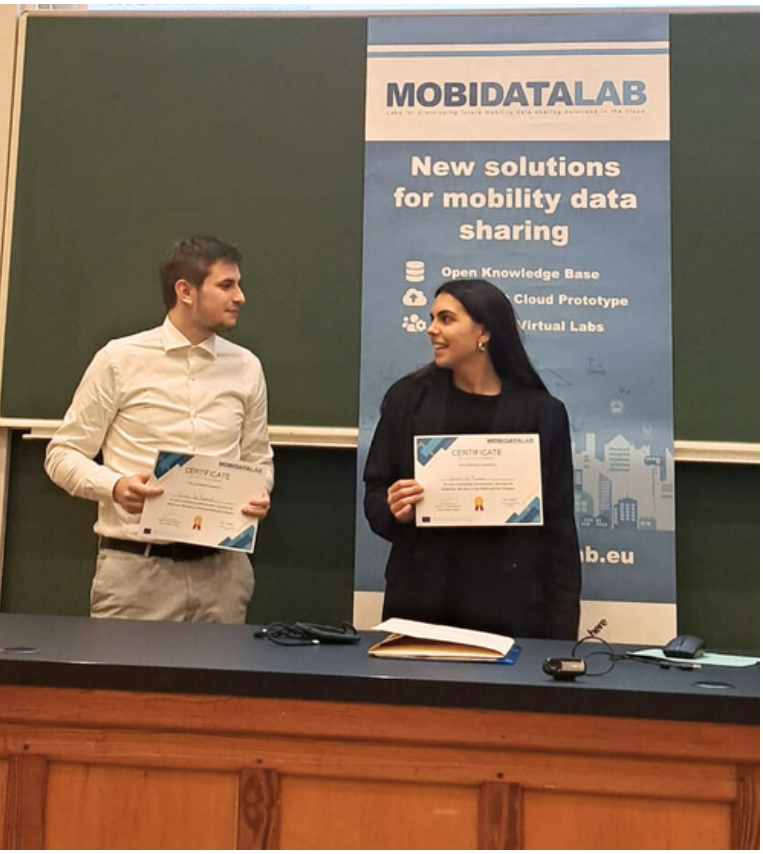
Unlike the traditional Datathons and Hackathons that often span just a couple of intense days, Codagon (stemming from the word ‘code’ and the Greek word ‘agon’ which stands for competition) offered a distinct approach to addressing mobility challenges, focusing on intensive collaboration and solution development over an abbreviated timeline of three weeks. Throughout its execution, the Codagon witnessed substantial participation, engagement, and impact, reflecting a concerted effort to address urban mobility challenges through innovation and collaboration.

With participation from over 98 individuals, including innovators, startups, researchers, and field experts, the Codagon fostered a diverse ecosystem of contributors. Spanning a duration of three weeks, from November 6th to November 27th, 2023, participants were provided with a platform to develop solutions to pressing urban mobility challenges. During this time, a series of strategic were implemented, including panel discussions, webinars, interactive sessions with challenge providers, and an onsite award ceremony in Leuven.

The success of the Codagon can be measured not only by the breadth of its engagement activities but also by the depth of collaboration and innovation it facilitated. Panel discussions, featuring expert panelists from diverse backgrounds, provided insights into cross-sector data fusion and opportunities for solution acceleration. Webinars, focusing on available technologies and resources, equipped participants with the tools and knowledge needed to develop robust solutions. Interactive sessions with challenge providers allowed for live feedback and clarification, enriching the solution development process.

The culmination of the Codagon, the onsite award ceremony held at the historic Emma Vorlat Auditorium at KUL, Leuven, Belgium, brought together finalists, challenge providers, mobility stakeholders, and domain experts. Here, finalists had the opportunity to showcase their solutions, receive feedback from stakeholders, and explore future collaborations. Notably, the synergies session emerged as a highlight of the event, fostering open dialogue and inclusivity among stakeholders.

In summary, the Codagon's impact transcended the confines of a conventional competition, leaving a lasting impression on urban mobility innovation. By fostering collaboration, facilitating knowledge exchange, and driving solution development, Codagon set a precedent for future initiatives in urban mobility, underscoring the importance of innovation and collaboration in addressing complex societal challenges.



3.3.4. The Virtual Lab

The Virtual Lab is a platform designed and validated within the MobiDataLab project with the aim to foster co-creation, problem-solving, and data-sharing in a collaborative environment primarily in the mobility realm.

The platform aims to solve the key needs of its two main user groups: organizations/municipalities and innovators (i.e. startups, developers, researchers, mobility experts). Organizations and municipalities are seeking impactful solutions to their complex problems and require robust data analytics for informed decision-making. They need a platform that can facilitate access to innovative solutions and streamline the process of addressing community and operational challenges through data-driven insights. Innovators, on the other hand, are looking for real-world applications for their ideas and technologies. They need access to challenges where they can pilot and test their solutions, collaborate with others, and receive feedback and recognition for their solutions. The Virtual Lab aims to meet these needs by providing a collaborative space for both groups, fostering innovation and problem-solving through community engagement and data sharing.

Within the MobiDataLab project, the Virtual Lab environment served as the primary tool for Living Labs participants to access challenge details, referenced datasets, and, most crucially, for the submission of solutions. The platform served also as the user interface of the MobiDataLab metadata catalogues directing participants to a plethora of transport and mobility datasets which served as valuable resources for their solution development.

This Virtual Lab environment provided a secure space for participants to upload their final solutions, ensuring a streamlined and standardised process for solution submissions. In response to the valuable feedback received during the first two MobiDataLab Living Labs, the Virtual Lab development team implemented various improvements to enhance the platform and turn it into its current MVP stage. MobiDataLab Datathon, Hackathon and Codagon serve as success stories of the platform's implementation, providing to more than 200 participants the opportunity to connect and showcase their innovative solutions in real-world mobility challenges provided by the cities of Paris, Milan, Eindhoven and Leuven.

Virtual Lab serves as a robust platform for municipalities, organizations, and innovators, aiming to accelerate innovation and transcend geographical boundaries for solving real-world mobility challenges.

4. MEASURING MOBILITY DATA-SHARING CULTURE

When delving into the intricate landscape of data-sharing culture within the transport sector, a robust methodology is paramount. In MobiDataLab, we proposed a comprehensive approach aimed at identifying claims and main topics derived from interviews conducted between June 2022 and September 2023. The essence of this methodology lies in its systematic framework, delineated into distinct stages to ensure a thorough analysis of the data-sharing ecosystem.

Our methodology begins with the establishment of an interview protocol. This protocol, meticulously crafted based on extensive literature analysis and aligned with the research objectives, lays the foundation for subsequent stages. The next step involves the identification of a diverse sample of interviewees. By encompassing various stakeholders within the transportation sector, ranging from policymakers to software providers, we ensured a holistic representation of perspectives.

Effective communication was pivotal in orchestrating successful interviews. The crux of our methodology lies in the meticulous execution of interviews. These sessions, recorded and transcribed, served as the bedrock for subsequent analysis and exploration. Leveraging cutting-edge argument mining tools, such as MARGOT, claims and evidence was extracted from the transcribed interviews.

In this stage, the groundwork was laid for discerning key assertions within the dataset. The textual corpus underwent rigorous preprocessing, which included tokenization, part-of-speech tagging, and entity recognition.

Advanced algorithms and tools were utilized to streamline the data for enhanced analysis. The culmination of our methodology was seen in the application of topic modeling techniques.

Through the use of Latent Dirichlet Allocation (LDA), underlying themes and topics embedded within the interview transcripts were unraveled, offering valuable insights into the landscape of data sharing. The amalgamation of Argumentation Mining and Topic Modelling formed the cornerstone of our analytical framework, facilitating a nuanced understanding of data-sharing dynamics within the transport sector. By highlighting claims and explaining main topics, our approach aims to clarify the complexities of this changing ecosystem.

Delving into the thematic realms uncovered through our methodology, several salient topics emerge, each encapsulating pivotal facets of the data-sharing culture within the transport sector:

1. ***Ensuring Data Quality and Completeness:*** *The pursuit of accuracy and thoroughness lies at the core of data-sharing, requiring a comprehensive approach to safeguard the reliability and integrity of shared datasets.*

2. ***Understanding Legal Responsibilities:*** *With evolving regulations, understanding the legal complexities of data-sharing is crucial, highlighting the importance of compliance and transparency.*

3 ***Building Collaboration and Trust:*** *Establishing strong collaborative frameworks built on trust is essential for fostering sustainable partnerships among diverse stakeholders.*

4 ***Sharing Previous Experiences and Knowledge:*** *Learning from past experiences and promoting transparency are key elements in cultivating a culture of knowledge-sharing and continual improvement.*

5 ***Considering Social and Environmental Impact:*** *Data-sharing initiatives have the potential to address environmental concerns and drive societal change, showcasing the transformative impact of data-driven solutions.*

6 ***Encouraging Data-sharing Motivations:*** *Effective communication and utilizing analytical capabilities play crucial roles in motivating data-sharing, enabling the achievement of common objectives.*

In essence, our methodological approach offers a comprehensive lens through which to dissect and comprehend the data-sharing culture within the transport sector. By unraveling key claims and topics, we aspire to contribute to a nuanced understanding of this dynamic ecosystem, paving the way for informed decision-making and transformative innovation.

5. BUILDING COMMUNITY: PROJECT'S OUTREACH

MobiDataLab embarked on a comprehensive outreach initiative aimed at fostering a robust and engaged community around data sharing and innovation in mobility.

Through strategic communication, dissemination efforts, and active engagement with various stakeholders, the project successfully built a vibrant ecosystem conducive to collaborative problem-solving and knowledge exchange. This summary highlights the key achievements and impactful activities undertaken by MobiDataLab to build and strengthen its community.

Central to MobiDataLab's community-building efforts was the development of the Virtual Lab environment. Notably, the MobiDataLab Datathon, Hackathon, and Codagon events served as prime examples of successful engagement, attracting over 200 participants who showcased innovative solutions to real-world mobility challenges.

MobiDataLab leveraged webinars and participation in external events as key channels for disseminating project results and engaging with target audiences.

Three webinars were organised [13][14] [15], each focusing on different aspects of data sharing and mobility improvement. These webinars not only provided valuable insights but also facilitated discussions and knowledge exchange among stakeholders. Additionally, the project actively participated in more than 20 external events, including conferences and workshops organised by both project partners and external organisations.

This extensive presence allowed MobiDataLab to reach a diverse audience and amplify its impact within the broader research and innovation landscape.

The project surpassed its initial collaboration targets by successfully engaging with more than 27 projects or initiatives. These collaborations spanned a wide range of domains, including research clusters, industry initiatives, and innovation incubators. By forging partnerships with such a diverse array of stakeholders, MobiDataLab demonstrated its proactive approach to community building and its commitment to fostering collaboration across different sectors and regions.

[13] https://www.youtube.com/watch?v=_Y3J5jEZN5w
[14] https://www.youtube.com/watch?v=BI9JBkFWteA
[15] https://www.youtube.com/watch?v=x9tDZmFhzzI&t=4s

In line with MobiDataLab's mission was the promotion of a data sharing culture within the mobility sector. Through its outreach activities, the project aimed to encourage mobility stakeholders to embrace open data principles and facilitate the exchange of data-driven insights and solutions. By organising events such as the Datathon and Hackathon, MobiDataLab provided tangible examples of how open data can drive innovation and address real-world challenges in mobility.

With partners located in seven countries, MobiDataLab demonstrated its commitment to fostering international collaboration and knowledge exchange. The project's outreach activities extended beyond the European Union, paving the way for the replicability of its approach and results in other regions. By engaging with a diverse range of stakeholders from different geographical locations, MobiDataLab enriched its community and contributed to the global conversation on data sharing and mobility.

MobiDataLab's journey towards building community has been one of connection, collaboration, and shared purpose. As we continue on this path, we remain committed to nurturing our community, driving positive change, and shaping the future of mobility together.

> *Overcoming stakeholders' reluctance was a challenge we addressed by showcasing the potential of data and fostering connections in the Living Labs. MobiDataLab has laid the groundwork for future mobility planning, and these achievements are stepping stones for continued progress.*
>
> ***Thierry Chevallier, Technical Project Coordinator, AKKODIS***

5. THE IMPACT OF MOBIDATALAB

Mobility data sharing has gained exceptional prominence, and expectations in the field have never been higher. If asked, both the consortium partners and the external stakeholders would be unanimous in saying that this interest has never diminished over the course of the project.

The remarkable attention MobiDataLab has attracted actually encouraged us to go beyond the objectives we had defined and exceed our initial expectations. MobiDataLab has therefore been an extraordinary experience for everyone involved, and now that the project is officially coming to an end, it offers tremendous opportunities for further development in a number of directions.

Firstly, MobiDataLab aimed to draw up a comprehensive inventory and assessment of the mechanisms promoting data sharing in the transport sector. As showcased throughout this document, dozens of projects, services and products were analysed in terms of their technical and commercial characteristics. The legal and governance aspects that represent one of the main obstacles to data sharing were studied in depth, with an analysis of the gaps in the selected legal frameworks and the availability of data in several EU countries.

Factors influencing data governance were also assessed (including applicable legal regimes such as GDPR), as well as mobility data protection techniques such as k-anonymity, micro-aggregation, encryption, etc. This inventory, which is being opened up to external contributors and therefore constituting now an "open knowledge base", also includes a study of the stakeholders' needs, as well as an overview of the business and revenue models that data sharing can enable. This knowledge base therefore provides a solid analysis of the current landscape for mobility data exchange, from a legal, technical, and economic perspective. The fact that it is based on authentic contributions from real representatives of the mobility sector, collected through our Advisory Board and Reference Group, significantly increases its value.

The impact of such a knowledge consolidation is already significant on the mobility data sharing ecosystem, and an essential asset for new projects currently being launched such as deployEMDS or NOUS. We expect this impact to continue in the years to come, through a joint exploitation plan consisting of continued content update, the adoption of a post-project governance with an open contribution approach, and finally the transfer of the knowledge base under an open license to a sustainable community association.

The second aim of the project was to prototype a scalable cloud-based solution for sharing mobility data – namely, the Transport Cloud – demonstrating the most effective ways of facilitating access to and exchange of mobility data. A reference metadata catalogue, compliant with data and service discovery standards such as DCAT-AP, has been deployed and now references over 100,000 datasets in the territorial context and specific domains of the MobiDataLab stakeholder "reference group". Data processors adding value to the data through several methods such as

the semantic and geographic enrichment were developed, and non-open datasets have been anonymised using innovative privacy preservation techniques which include information on confidentiality guarantees and re-identification risks. By agreeing to keep the platform online with demonstrators and to further develop open-source components such as the geo-semantic enrichment processor, the MobiDataLab partners are now able to involve a wider range of data providers and enable more advanced collaboration between transport operators for the co-development of multimodal solutions. The platform's alignment with the data space architecture is a realistic technical possibility in the very near future, given the major developments underway in this ecosystem, and it is an extraordinary opportunity to deliver a mobility data space augmented with services tailored to solving concrete urban mobility problems.

Finally, the third aim of MobiDataLab was to gather representatives of data providers and data consumers and conduct agile exchange sessions aiming to co-create, explore, experiment, and evaluate the Transport Cloud – namely, the Living and Virtual (data) Labs. During the creation of the knowledge base and the prototyping of the Transport Cloud, relevant use cases were identified, allowing the definition of new requirements improving the findability, accessibility, interoperability, and reusability of mobility data. These use cases have been contextualised and refined by the MobiDataLab Reference Group of Mobility stakeholders, leading to so-called “challenges” proposed to data engineers and experts such as multimodal mobility hubs’ location in Leuven, pedestrian mobility improvement in Paris, taxi mobility analysis in Eindhoven, or large events accessibility in Milan. The three MobiDataLab innovation events organised – the Datathon, the Hackathon, and the Codagon – have been a real success and have had a significant impact on a community that has expanded as a result. In addition, the 200 individual registrations received offer promising scaling opportunities, as the project created a significantly large community of mobility data users that transport stakeholders seek to create and maintain on their side. By maintaining the virtual lab online with updated challenges, reinforcing the collaboration between the talented participants and their challenge provider, and providing guidance on how to replicate the innovation sessions in other contexts, local areas, cities, regions, etc., MobiDataLab will continue to have a significant impact on the mobility data ecosystem, strengthening the data sharing culture through this unique methodology.

In closing, MobiDataLab leaves behind a legacy of remarkable achievements and significant potential for further advancement. The project's impact has been profound, not only in fostering collaboration and innovation but also in shaping the future of mobility data exchange. With its open knowledge base, transformative Transport Cloud solution, and dynamic Living and Virtual Labs, MobiDataLab has set the stage for a future where data sharing in the transport sector is more accessible and impactful than ever before. As we transition to the next phase in this digital era, we eagerly anticipate the continued growth and evolution of the mobility data ecosystem, fuelled by the insights and experiences gained through MobiDataLab.

MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

The 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

https://mobidatalab.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101006879. The content of this article reflects solely the views of its authors.