MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

D5.3 Analysis and Conclusions on the Data Exchange Culture

01/02/2024

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	D5.3
Deliverable Name	D5.3 - Analysis and Conclusions on the Data Exchange Culture
Full Project Title	MobiDataLab, Labs for prototyping future Mobility Data-sharing cloud solutions
Responsible Author(s)	Selini HADJIDIMITRIOU (ICOOR)
Contributing Partner(s)	AETHON, POLIS, HERE, AKKA
Peer Review	F6S, HOVE
Contractual Delivery Date	31-01-2024
Actual Delivery Date	30-01-2024
Status	Final
Dissemination level	Public
Version	V1.0
No. of Pages	44
WP/Task related to the deliverable	WP5/T5.2
WP/Task responsible	ICOOR
Document ID	MobiDataLab-D5.3-AnalysisAndConclusionsOnTheDataExchangeCulture_v1.0
Abstract	This deliverable describes the final analysis of the data-sharing culture.

Legal Disclaimer

MOBIDATALAB (Grant Agreement No 101006879) is a Research and Innovation Actions project funded by the EU Framework Programme for Research and Innovation Horizon 2020. This document contains information on MOBIDATALAB's core activities, findings, and outcomes. The content of this publication is the sole responsibility of the MOBIDATALAB consortium and cannot be considered to reflect the views of the European Commission.

Project partners

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

Document history

Version	Date	Organization	The main area of changes	Comments
0.1	21/01/2024	ICOOR	All	Creating the first draft
0.2	23/01/2024	HOVE, F6S	Whole document	Review
0.3	25/01/2024	ICOOR	Whole document	Implementation of reviews
0.4	25-26/01/2024	ICOOR, AKKODIS	Whole document	TL & Coordinator QC
1.0	30/01/2024	AKKODIS	Whole document	Final Version & Submission

Funded by the European Union

Executive Summary

This deliverable presents the results of the Transport Cloud acceptance and data-sharing culture evaluation. The assessment of the Transport Cloud's acceptance through the MobiDataLab project reveals valuable insights into its effectiveness, user perceptions, and areas for improvement. The transition from stakeholder expectations to practical experiences during the x-athons, involving datathon, hackathon, and codagon events, provides a comprehensive view of the platform's evolution. The evaluation of the Transport Cloud reveals a dynamic landscape of user perceptions, emphasizing the need for continuous engagement, clear communication, and iterative improvements.

Concerning the data-sharing culture, data quality and completeness in the context of mobility underscore the intricate interplay between accuracy, representativity, and the open availability of data. Ensuring accurate data is a starting point, but anonymization without compromising quality is crucial. Advanced data science skills are required to navigate potential pitfalls and make datasets meaningful for transportation analysis. Direct connections with data owners, awareness of limitations, and clear expectations are vital for improving overall data quality. Legal responsibility in data-sharing for public services involves evolving regulations like GDPR. The collaboration between the public and private sectors emphasizes transparency, personal data control, and a shared goal of an interconnected and transparent transport system. Knowing how data is used is crucial for building public trust. Establishing trust is key in collaborations, especially with diverse stakeholders. Governmental authorities play a crucial role in building and maintaining trust. Trust can be strained when partners do not respect research questions or provide low-quality data. Partner selection involves balancing fairness, encouraging participation, and diversifying opportunities.

Previous experience and knowledge-sharing are crucial for creating long-term value in data-sharing. Transparency about successes and challenges, adapting to evolving realities, and learning from past experiences contribute to a culture of collaboration. EU projects provide motivation, fostering commitment and satisfaction among participants.

In the pursuit of sustainable transportation, data-sharing contributes to optimizing logistics and reducing environmental impact. Demonstrating the benefits of public transport through data promotes environmental awareness. Machine learning plays a role in recommending environmentally friendly transport modes. Rigorous measurement of environmental impact through data analytics is essential for achieving sustainability goals.

Sharing data and tools is essential for achieving common goals and building necessary infrastructure. Effective communication and information exchange are vital for promoting the positive aspects of data-sharing. Encouraging experts to work in public institutions is crucial for leveraging analytical capabilities in public mobility planning.

Table of content

Tables

Figures

Funded by the European Union

Abbreviations and acronyms

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

Funded by the European Union

1. Introduction

This deliverable provides a comprehensive overview of the final evaluation results of the data-sharing culture and Transport Cloud acceptance. Specifically, D5.3 compares the pre and post-survey results on Transport Cloud acceptance and analyzes 40 interviews with actors of the data-sharing culture.

The survey on Transport Cloud acceptance adheres to the Unified Theory of Acceptance and Use of Technology (UTAUT), emphasizing key factors such as performance expectancy, effort expectancy, and facilitating conditions. The UTAUT framework in the survey design demonstrates a structured and well-established approach to understanding technology acceptance. This framework, encompassing performance expectancy, effort expectancy, and facilitating conditions, provides a comprehensive lens for evaluating user perceptions and behaviors. The delineation of the survey into macro areas aligned with the UTAUT framework allows for a systematic examination of critical aspects influencing technology acceptance.

The comparative analysis between pre and post-survey results offers insights into the evolution of user perceptions over time. Identifying shifts or consistencies in responses provides valuable information about the impact of the Transport Cloud and the need for adjustments. Evaluating performance expectancy helps gauge users' perceptions of the value and benefits offered by the Transport Cloud. This aspect is fundamental in understanding the platform's utility in meeting user needs and expectations. Assessing effort expectancy provides insights into users' perceptions of the ease of use and accessibility of the Transport Cloud. User-friendly interfaces and seamless interactions contribute to positive technology acceptance. Understanding the influence of expectations is essential for addressing broader contextual factors affecting technology acceptance. Furthermore, examining facilitating conditions helps evaluate the supporting infrastructure and resources that enable or hinder the adoption of the Transport Cloud.

Concerning the evaluation of the data-sharing culture, the inclusion of 40 interviews adds qualitative depth to the evaluation process. Most importantly, engaging with a diverse set of actors allows for a more nuanced understanding of their perspectives, challenges, and experiences related to data-sharing.

1.1. Purpose of the document and target group

The overarching objective of this deliverable is to conduct a comprehensive evaluation of the data-sharing culture, specifically focusing on assessing the impact of the Transport Cloud in enhancing data-sharing.

Through the analysis of survey results and in-depth interviews with various actors, the goal is to gain insights into whether the Transport Cloud prototype developed within the MobiDataLab project catalyzes promoting the exchange of data. Furthermore, the focus of the evaluation is on understanding the dynamics of data-sharing culture.

Funded by the European Union

The interviews with various actors in the data-sharing culture provide qualitative insights into the key factors that influence data exchange. By engaging with diverse stakeholders, the evaluation aims to uncover perspectives and identify elements that contribute to or impede a collaborative data-sharing environment. Furthermore, the evaluation seeks to uncover the critical factors that stimulate data exchange among users. It delves into the motivations, challenges, and facilitating conditions that shape the culture of data-sharing in the context of the Transport Cloud.

Through a holistic analysis of survey responses and interview results, the evaluation aims to provide actionable insights to improve the culture of data-sharing. This may involve recommendations to refine the Transport Cloud prototype, address user concerns, or implement strategies to further encourage data-sharing.

The deliverable seeks to translate the results into practical implications for the continued development and optimization of the Transport Cloud. The goal is to guide future efforts to align the platform with user expectations and foster a collaborative ecosystem for data-sharing.

In summary, the deliverable is a comprehensive exploration of the acceptance of the Transport Cloud and an analysis of the culture of data-sharing. Through a combination of quantitative survey analysis and qualitative stakeholder interviews, the assessment aims to uncover the complexities of user experiences and provide valuable insights to optimize the platform and promote a robust data-sharing environment.

1.2. Contribution of partners

The substantial contribution of the MobiDataLab participants played a key role in shaping the content and results of this deliverable. The key was the involvement of the participants in both the evaluation of the Transport Cloud prototype and the extensive interview process.

The organization of events, during which the organizers presented the Transport Cloud prototype and participants used the Virtual Labs, created the context for participant involvement. Participants at these events filled out the survey and this active participation provided first-hand information on users' perspectives on Transport Cloud. The MobiDataLab partners' extensive contribution to sharing the first phase of the survey is an integral part of the evaluation process.

Their engagement reflects a collaborative effort within the consortium and ensures a diverse contribution from stakeholders to the success of the Transport Cloud.

The wide network of stakeholders within the consortium played a key role in reaching out to a diverse range of actors for the interviews. The consortium's wide network facilitated the inclusion of different perspectives, enriching the qualitative insights gathered through the interview process. Interviewees included MobiDataLab partners, demonstrating their commitment and involvement. The partners' views, as contributors and users, offer valuable insights into the platform's impact and areas for improvement.

Survey data, partner contributions, and interviews contribute to a holistic evaluation approach. This approach is essential for understanding user experiences and the effectiveness of the platform.

The collaborative nature of the MobiDataLab consortium is evident in the collective effort to contribute to the outcome. The variety of participants and their contributions highlights the strength of collaboration within the consortium.

1.3. Relation to other activities

The final version of Deliverable D5.3 serves to conclude D5.2, thus finalizing the activities of T5.2. This continuity underlines the progression of the evaluation process from the preliminary investigation and analysis of the data-sharing culture, as initially outlined in D5.2. Furthermore, the methodology, initially outlined in D3.5, defines the natural progression of the evaluation process and a natural relationship between D5.3 and WP3.

The activities outlined in D5.3 represent a logical progression from the preliminary survey and data-sharing culture analysis presented in D5.2. The continuity of the methodology ensures a coherent and structured approach to the Transport Cloud impact assessment and data-sharing culture analysis. The iterative nature of the evaluation process is captured in the transition from initial assessments to the final version, providing a comprehensive understanding of the impact of the Transport Cloud and the culture of data-sharing.

Funded by the European Union

2. Evaluation of the Transport Cloud

The assessment and analysis of the impact of the Transport Cloud are crucial steps in understanding its effectiveness and gathering insights for further improvements. The comparison between the pre-Living Labs demonstration survey and the responses collected during the MobiDataLab events (x-athons) provides valuable perspectives on how perceptions and experiences have evolved.

The x-athons, comprising the datathon, hackathon, and codagon, served as practical environments for MobiDataLab partners to interact with and test the Transport Cloud and Virtual Lab functionalities. The differentiation in participant numbers between the two survey rounds (39 in the first round and 19 in the second round) may be attributed to the shift from stakeholders' expectations assessment to actual user experiences during the x-athons.

The initial survey aimed at stakeholders involved in the MobiDataLab project sought to gauge expectations for the Transport Cloud before its implementation. Some respondents found it challenging to answer certain questions without the ability to test a Transport Cloud prototype. This highlights the inherent difficulty in predicting user experiences and expectations in the absence of hands-on interaction.

In the second round, participants in the x-athons had a concrete experience with the Virtual Labs, which integrated the Transport Cloud. The presentation of the Transport Cloud, including CKAN and GeoNetwork, along with the additional functionalities of the Virtual Lab, set the stage for participants to engage with datasets, services catalogs, challenges, and solution submissions.

To facilitate a clear comparison, responses are reported in percentages. This approach allows for a relative assessment, considering the variation in participant numbers between the two survey rounds.

The transition from expectations to actual experiences provides a comprehensive view of how the Transport Cloud is perceived in practical usage scenarios. Feedback gathered during the x-athons, where participants engaged with real datasets and explored additional functionalities, offers valuable insights into the platform's strengths, weaknesses, and areas for improvement.

As the assessment continues, the data collected from these surveys becomes instrumental in refining the Transport Cloud and shaping its future implementations. The ongoing feedback loop, especially from participants who have interacted with the platform, is crucial for iterative enhancements and ensuring that the Transport Cloud aligns with user needs and expectations.

Funded by the European Union

2.1. Type of respondents in the prep and post-survey

The change in the composition of participants from the preliminary (pre) survey to the x-thons indicates a strategic move towards a more specialized and targeted audience for Transport Cloud/Virtual Labs. This change reflects a deliberate effort to engage people with specific skills and interests in data science techniques.

In the preliminary survey, the group of participants showed greater heterogeneity, displaying a diverse range of stakeholders with different backgrounds and interests. This heterogeneity included individuals from different sectors and roles, each of whom brought unique perspectives to the evaluation of the Transport Cloud (Figure 1).

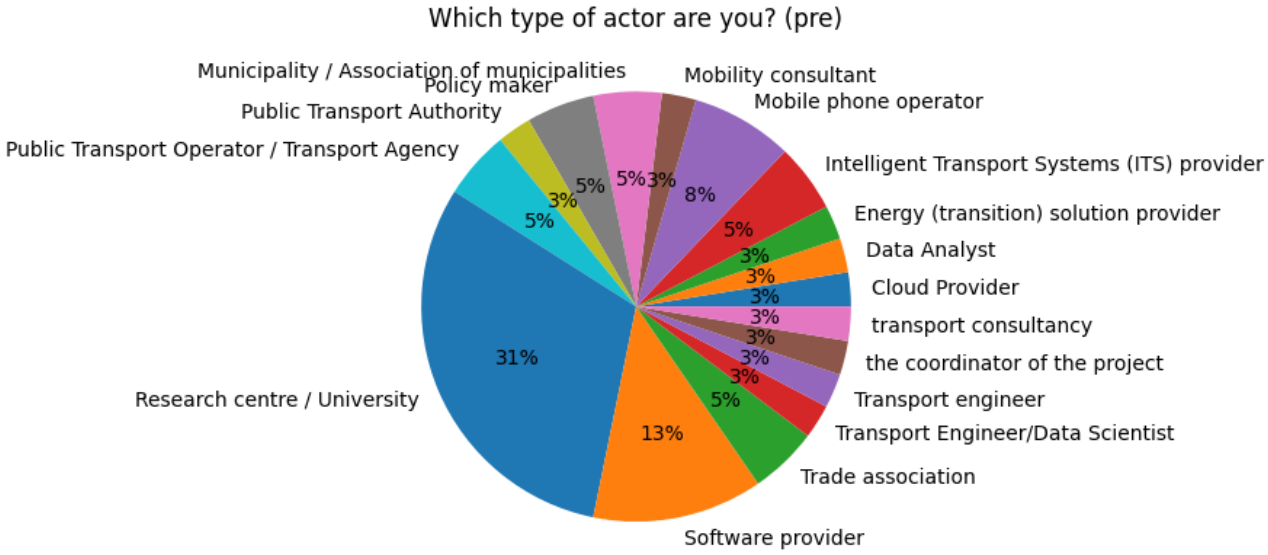


Figure 1: Type of actor (pre-survey)

The explanation provided for the higher heterogeneity in the preliminary survey aligns with the inherent nature of the Transport Cloud/Virtual Labs. These platforms are designed to cater to individuals well-versed in data science techniques. As such, the functionalities and datasets offered by the Transport Cloud may be most relevant and appealing to a more specialized group of users who possess the expertise required for effective data handling and analysis.

By converging towards a more specialized audience in the X-thons, the project organizers may have sought to gather feedback and insights from users who can fully leverage the advanced features of the Transport Cloud. This targeted approach allows for a deeper exploration of the platform's capabilities and a more nuanced understanding of its effectiveness in addressing the specific needs of data science professionals.

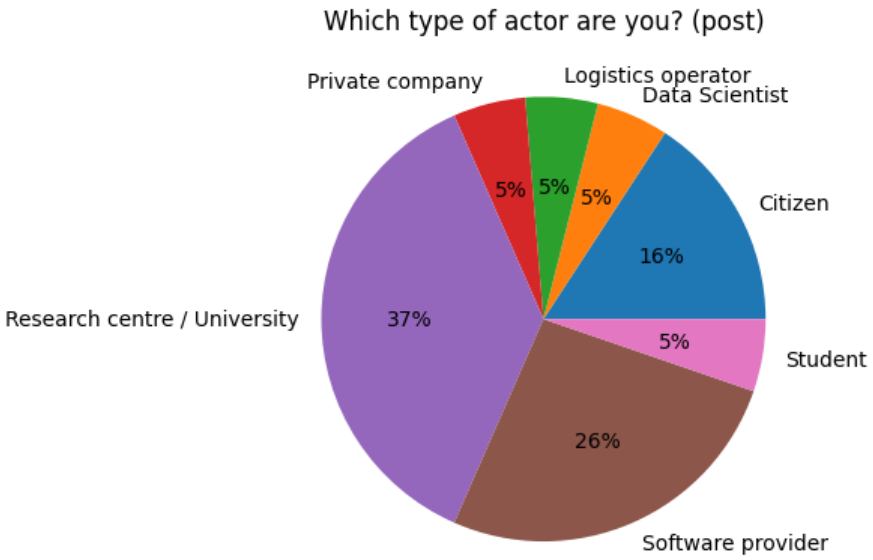


Figure 2: Type of actor (post survey)

The shift in participant composition observed in the x-thons, with a majority from Research centers/Universities (37%) and Software providers (26%), is noteworthy and aligns closely with the participant distribution seen in the preliminary survey. This consistency indicates a continuity in the engagement of specific actor types, emphasizing the importance of involving research and software professionals in the assessment of the Transport Cloud/Virtual Labs (Figure 2).

In both the preliminary survey and the x-thons, Research centers/Universities represented a significant portion of the participants, underscoring the academic interest and relevance of the Transport Cloud in research environments. These institutions often play a crucial role in exploring and adopting innovative technologies, making them valuable contributors to the evaluation process.

The presence of Software providers in both the preliminary survey and the x-thons is also consistent, albeit with a higher percentage in the x-thons. This suggests a continued interest and involvement of professionals from the software industry in assessing the Transport Cloud/Virtual Labs. Their participation is crucial for evaluating the platform's technical aspects and aligning it with industry standards and requirements.

The notable shift in the x-thons towards a composition closely resembling that of the preliminary survey indicates a deliberate effort to maintain engagement with these specific actor types.

It may reflect a targeted approach to ensure continuity in gathering insights from participants who are likely to have a deep understanding of data science, software development, and related domains.

Funded by the European Union

Overall, this consistent composition across both phases of the project suggests a deliberate strategy in involving actors from Research centers/Universities and Software providers. This approach ensures a sustained focus on academic and industry perspectives, contributing to a comprehensive evaluation of the Transport Cloud/Virtual Labs in both research and practical application contexts.

The consistency in the representation of these two dominant actor types across both surveys suggests that the Transport Cloud/Virtual Labs are resonating particularly well with these specific groups. Research centers and universities often have a strong focus on data-driven research, making them natural users of platforms that facilitate data access and analysis. Software providers, on the other hand, likely find value in the technical aspects and functionalities of the Transport Cloud.

While the shift towards a more focused audience may limit the diversity of participant types, it also indicates that the Transport Cloud/Virtual Labs are effectively meeting the needs of those directly involved in research, software development, and related fields. This alignment with the platform's target users enhances the likelihood of gathering meaningful feedback and insights from individuals who can provide valuable perspectives on its capabilities and potential improvements.

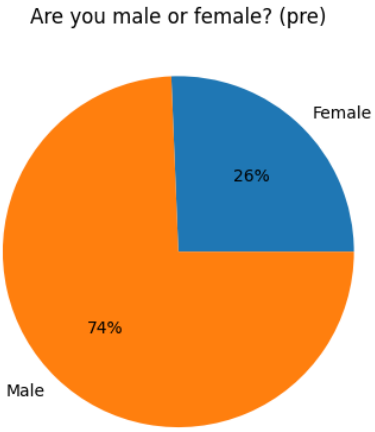


Figure 3: Gender (pre survey)

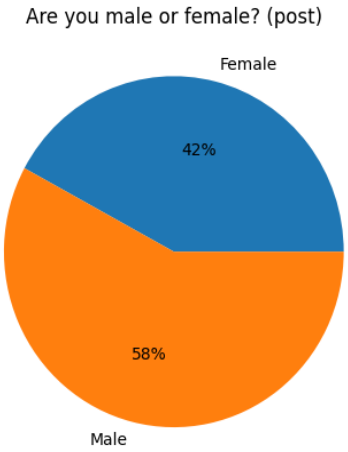


Figure 4: Gender (post survey)

The observed change in gender balance between the preliminary survey and the final round of interviews suggests a positive shift towards a more balanced participation, with an increased representation of females.

In the preliminary survey, the gender distribution was skewed, with 74% of participants being male. However, in the final round of interviews, this balance improved, with the percentage of males decreasing to 58%. This change indicates a more inclusive engagement of individuals from different gender backgrounds in the later stages of the project.

The increased representation of females in the final round could be attributed to various factors. It may reflect intentional efforts to reach a more diverse audience in subsequent survey rounds or events. Alternatively, the nature of the final round, such as specific outreach strategies or event structures, may have naturally attracted a more diverse group of participants.

A balanced gender representation is crucial for obtaining a comprehensive understanding of user experiences and preferences. It contributes to a more inclusive and representative assessment of the Transport Cloud/Virtual Labs, ensuring that feedback and insights gathered are reflective of a broader user demographic.

As projects evolve, recognizing and addressing any initial gender imbalances can be important for fostering diversity and inclusivity. The positive shift observed in the final round signals progress towards a more equitable participation in the evaluation and feedback processes related to the Transport Cloud initiative.

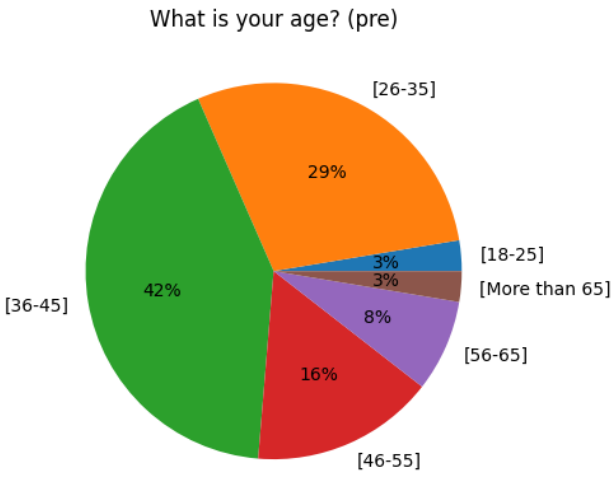


Figure 5: Age (pre survey)

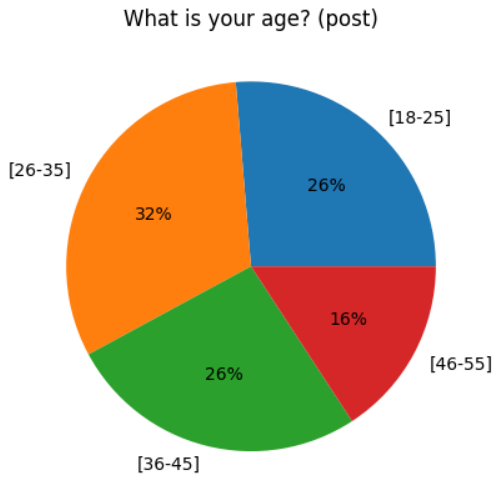


Figure 6: Age (post survey)

The increased participation of students in the x-thons, with a representation of 26% in the final survey compared to only 3% in the preliminary survey, highlights a deliberate effort to engage a younger demographic in the evaluation of the Transport Cloud/Virtual Labs.

The shift toward a more satisfactory representation of young participants aligns with contemporary trends, acknowledging the significance of involving the younger generation, particularly students, in projects and initiatives related to technology, data science, and cloud computing.

The overall good representativeness of the age range between 26 and 55 in both the preliminary and final surveys suggest a well-balanced engagement across a broad spectrum of age groups. This diversity in age representation is essential for capturing a range of perspectives, experiences, and expectations, contributing to a comprehensive understanding of how the Transport Cloud/Virtual Labs are perceived across different life stages and professional experiences.

Including students in the evaluation process not only enriches the feedback received but also provides an opportunity to assess the platform's usability and appeal to those who may be newer to the workforce or academia. It reflects a forward-thinking approach to technology adoption, recognizing the importance of preparing the next generation of professionals with experience in cutting-edge tools and platforms.

2.2. Transport Cloud acceptance

The survey on the Transport Cloud, designed to evaluate the tool's level of acceptance among different actors, employs the acceptance model outlined in D5.2. This model encompasses various factors such as usability, functionality, and perceived benefits. The first crucial question posed to survey participants is whether the Transport Cloud brings tangible benefits to them.

This initial question addresses the fundamental aspect of user satisfaction and utility. By directly asking participants whether they perceive benefits from using the Transport Cloud, the survey aims to gauge the tool's positive impact on their workflows, projects, or specific needs.

The responses to this question provide valuable insights into the real-world utility of the Transport Cloud, offering a qualitative measure of its effectiveness from the users' perspective. Positive responses would indicate that the tool aligns with user expectations and contributes positively to their tasks, while negative or ambivalent responses could highlight areas for improvement.

By framing the survey with this foundational question, the project team gained a nuanced understanding of how the Transport Cloud is received by its user base. Subsequent analysis of responses can inform strategic decisions, user engagement strategies, and potential enhancements to optimize the tool's acceptance and utility.

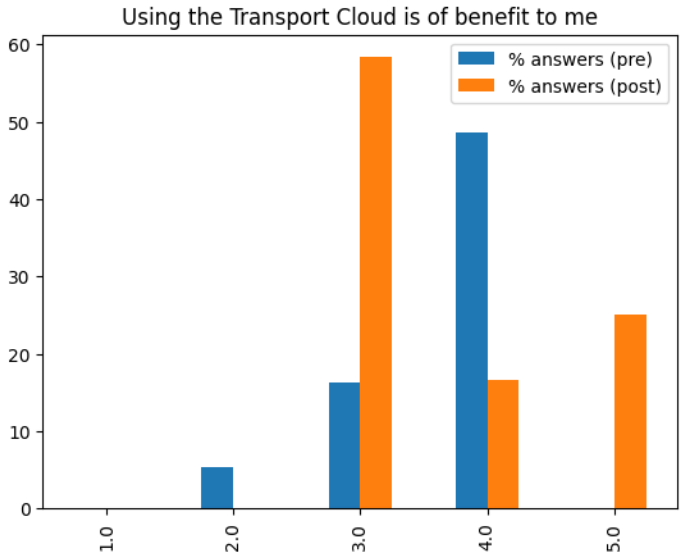


Figure 7: Benefits of using the Transport Cloud

The shift in average scores from the preliminary survey to the final one, indicating a decrease in perceived usefulness, is a significant finding that warrants careful consideration. The decrease from an average score of 4 (49%) in the preliminary survey to 3 (58%) in the final survey suggests a shift in participant perceptions, with those in the X-thons considering the Transport Cloud less beneficial.

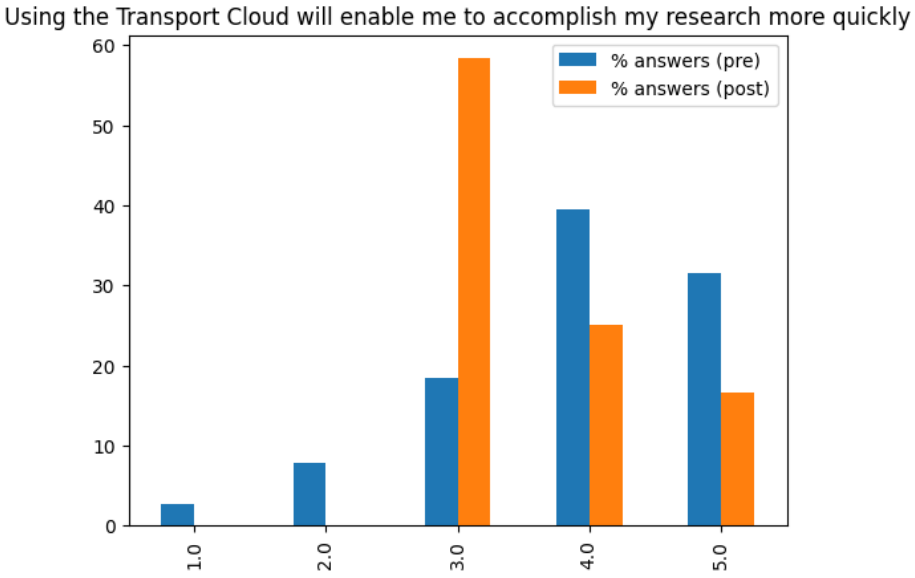


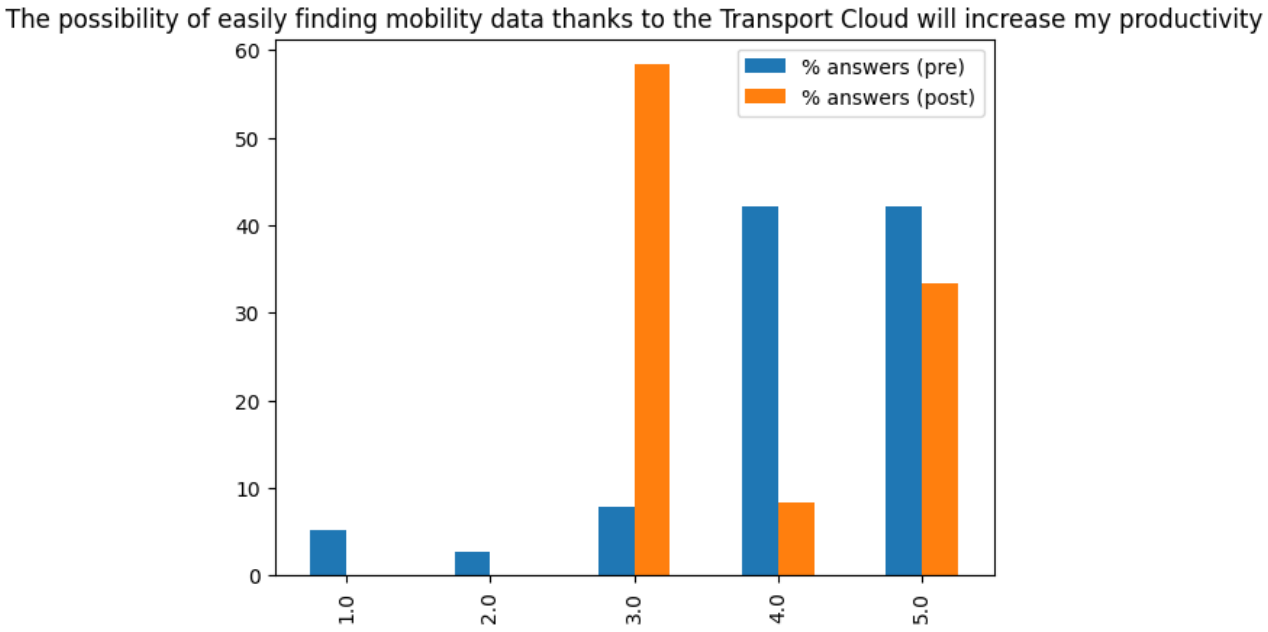
Figure 8: Usage of the Transport Cloud to accomplish research activities more quickly

The observed decrease in scores related to the Transport Cloud's acceptance, particularly in terms of aiding in accomplishing research results more quickly, is a noteworthy trend. This shift indicates a potential misalignment between the initial expectations set at the beginning of the project and the actual experiences of participants, with a tendency for participants to rate the tool-less favorably over time.

The decrease in scores suggests that initial expectations about the Transport Cloud might have been set at a higher level than what participants experienced in practice. This highlights the importance of managing expectations effectively, ensuring that users have a realistic understanding of the tool's capabilities.

The difference in scores between software providers and researchers/logistics operators is a crucial aspect. The higher scores from software providers indicate that they perceive the Transport Cloud as more useful. Understanding the reasons behind this divergence can provide valuable insights into how different user groups interact with the tool. Furthermore, the finding that software providers found the Transport Cloud more useful than researchers or logistics **operators** points to potential discrepancies in the tool's alignment with specific user needs. Exploring the features or functionalities that resonate with software providers can guide efforts to enhance the tool's relevance for other user groups.

The positive shift in participants' agreement regarding the Transport Cloud's capability to improve access to open data, with the average increasing from three to four and five in the second round of interviews, is a notable success for the MobiDataLab project. This outcome aligns with one of the project's key objectives: to prototype a platform that enhances the searching, accessing, and



fusing of mobility data in the cloud.

Figure 9: The Transport Cloud increases the productivity

The increased agreement among participants regarding the Transport Cloud's ability to improve access to open data directly aligns with the MobiDataLab project's goals. This suggests that the platform is effectively fulfilling its intended purpose of providing enhanced accessibility to mobility data. The increased average agreement implies that participants recognize the value of the Transport Cloud in facilitating easier access to open data.

This is particularly important for a project focused on mobility data, where data accessibility is a key factor in deriving meaningful insights. The mention of "fusing mobility data in the cloud" indicates that the Transport Cloud is contributing not only to data access but also to data integration and fusion. This multifaceted functionality enhances its overall utility for users involved in mobility data analysis.

The positive response to the Transport Cloud's impact on open data access can be seen as a broader indicator of the project's success. If the platform is effectively meeting user expectations in this crucial aspect, it bodes well for the overall success of the MobiDataLab initiative. Understanding that users perceive the Transport Cloud positively in terms of data access can guide future development efforts.

It provides insights into what aspects of the platform are working well and can inform strategies for further enhancements.

2.3. Easiness of use of Transport Cloud

The notable improvement in participants' perceptions regarding the ease of becoming skillful in using the Transport Cloud is a positive indication of the project's progress and the platform's user-friendliness. According to the acceptance theory employed for the survey questions, this improvement can be a key predictor of overall acceptance of the Transport Cloud.

The shift from 30% of respondents considering it very easy to become skillful in using the Transport Cloud at the beginning of the project to 70% in the latest responses signifies a positive change in participant perceptions. This shift suggests that the platform has likely undergone improvements in terms of user onboarding, training, or overall usability. According to the acceptance theory, the perceived ease of use is often a strong predictor of overall system acceptance. The substantial increase in respondents finding it easy or very easy to become skillful indicates a positive correlation with the anticipated acceptance of the Transport Cloud. The project's educational efforts, such as training sessions, documentation, or support mechanisms, may have contributed to the increased confidence and perceived ease of becoming skillful. This highlights the importance of ongoing educational initiatives in the successful adoption of technological tools.

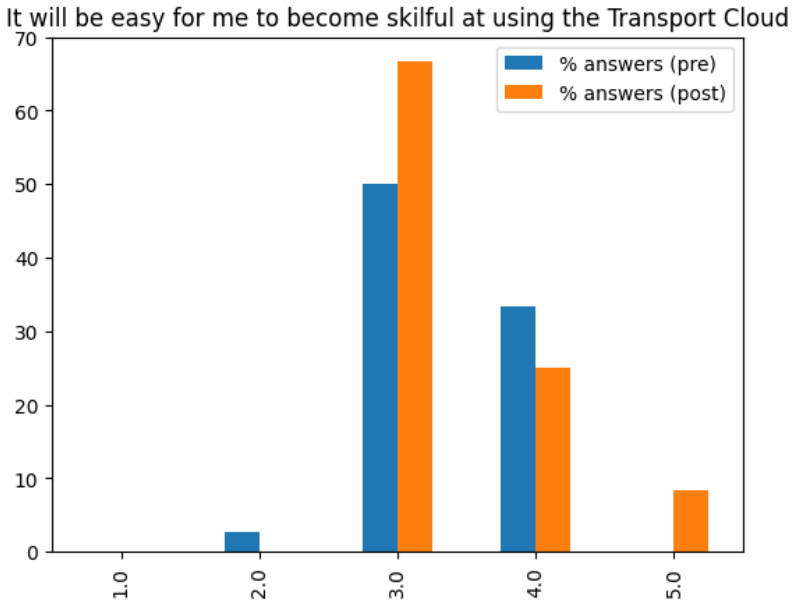


Figure 10: Easy to become skillful in using the Transport Cloud

The observed shift in the distribution of answers on the easiness of using the Transport Cloud between the pre-survey and post-interviews, with the post-interviews showing a more concentrated distribution around a medium agreement score of three, is a common and expected phenomenon in the context of technology adoption and user familiarity.

Participants' initial uncertainty in the pre-survey may be attributed to the novelty of the Transport Cloud and the learning curve associated with any new technology. With hands-on experience, users may have gained a more realistic understanding of the tool's ease of use.

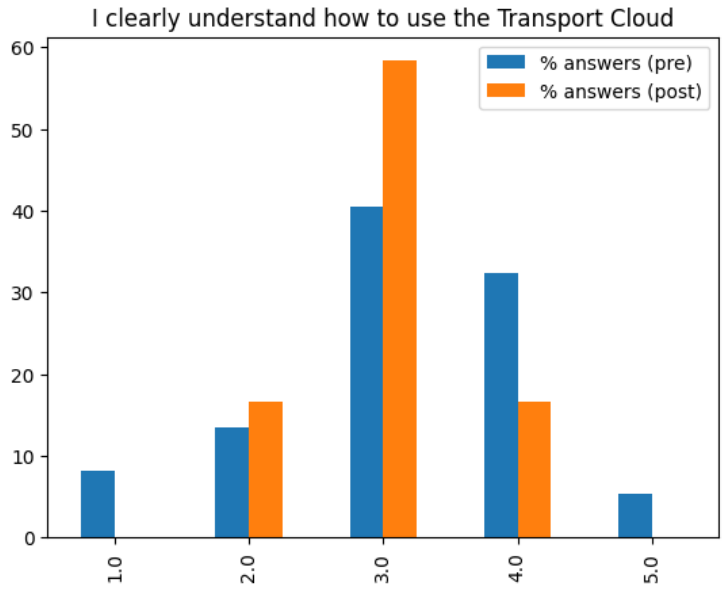


Figure 11: Easy to understand how to use the Transport Cloud

The heterogeneity in answers in the pre-survey is understandable, especially if participants had not yet interacted with the Transport Cloud prototype. Users often find it challenging to assess the ease of using a tool they haven't experienced first-hand. As participants progressed through the project and engaged with the Transport Cloud prototype during the post-interviews, their understanding of its usability likely became clearer. The shift towards a more concentrated distribution around a medium agreement score suggests a more informed assessment.

2.4. Transport Cloud and Data Security

The observed increase in confidence in the Transport Cloud's capability to ensure data security, as reflected in the post-interview responses compared to the preliminary survey, is a positive development, particularly in a sector where sensitive information, such as location data, plays a fundamental role in understanding traveler behavior.

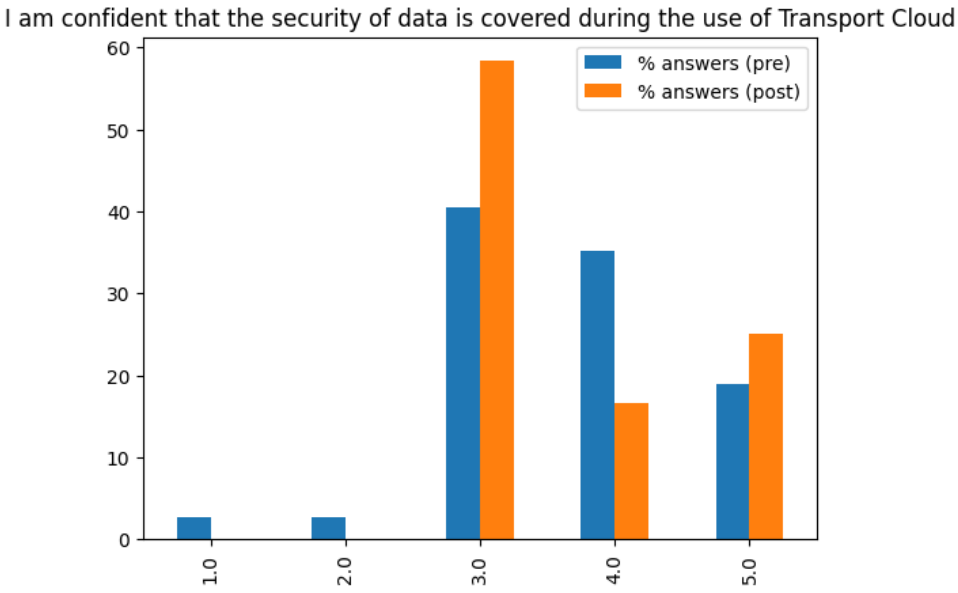


Figure 12: Security of data is covered during the use of the Transport Cloud

The increased confidence in the Transport Cloud's capability to ensure data security, especially in the context of privacy-preserving and security-enabling techniques, can indeed be attributed to MobiDataLab's proactive approach to addressing privacy and security issues associated with data-sharing. The fact that one of the project objectives explicitly dealt with identifying privacy and security issues associated with data-sharing underscores the project's commitment to understanding and addressing these concerns. Participants may have recognized the project's diligence in addressing potential security challenges.

The MobiDataLab's consideration of privacy-preserving and security-enabling techniques from different perspectives indicates a comprehensive approach to safeguarding mobility data. This comprehensive perspective likely instilled confidence among participants that the Transport Cloud was designed with a robust security framework.

2.5. Transport Cloud data-sharing culture and satisfaction

The evolution of perceptions regarding the Transport Cloud and its impact on data-sharing culture is a fascinating journey. The initial findings, where only 45% believed in its potential, highlight the challenges faced in introducing a new concept. However, the remarkable shift to 90% satisfaction among respondents demonstrates a significant success in changing perceptions and fostering a positive data-sharing culture.

The initial lack of clarity about the Transport Cloud concept is a common challenge in the early stages of introducing innovative technologies. The transition from a limited understanding at the project's onset to a more comprehensive grasp in the post-survey indicates successful efforts in communicating and educating participants about the platform.

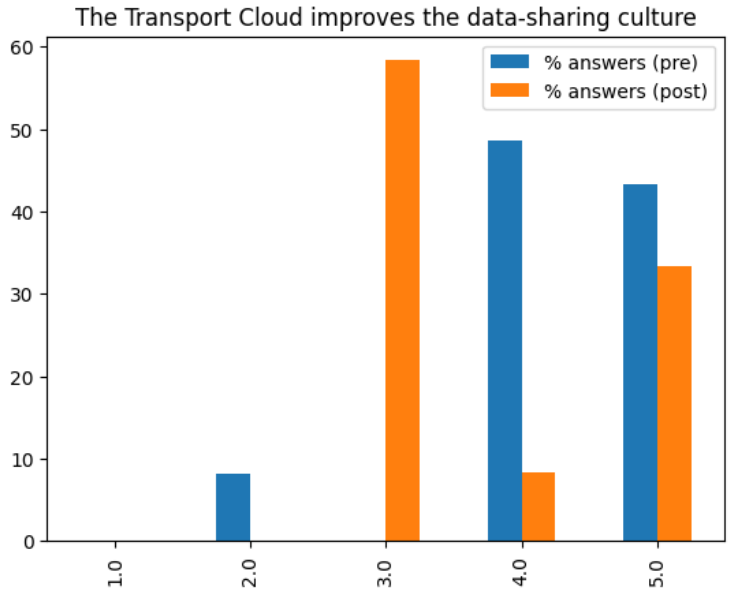


Figure 13: The Transport Cloud improves the data-sharing culture

While the overall satisfaction with the Transport Cloud's impact on data-sharing culture is high, the nuanced decrease in satisfaction with the Transport Cloud concept in the post-survey suggests that there might be unmet expectations or areas for improvement. The shift from a satisfaction level of four to three for some participants may indicate that their expectations were slightly higher than what was delivered.

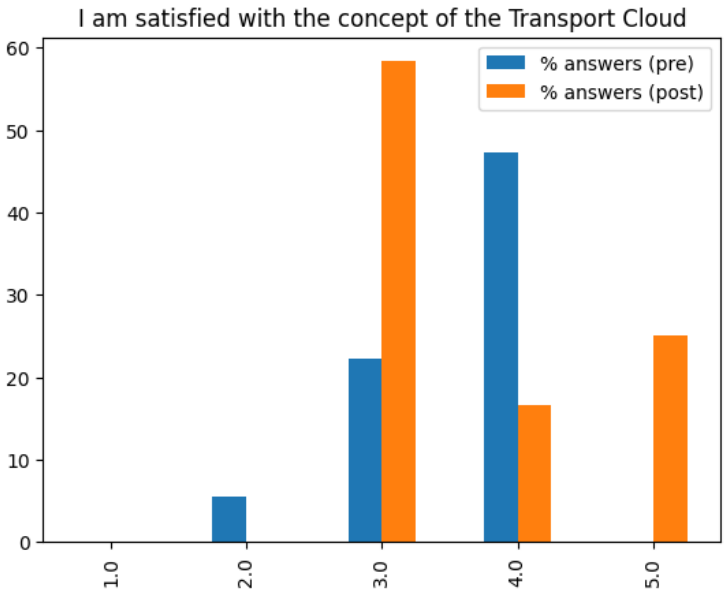


Figure 14: Satisfaction with the Transport Cloud concept

The absence of open-ended questions to delve into the reasons behind the satisfaction levels limits the ability to pinpoint specific areas of concern.

However, the mention of expectations being slightly higher suggests a need for clearer communication, ongoing education, or potential enhancements to meet or exceed user expectations.

The exciting prospect of further extending the Virtual Lab with new functionalities and concepts using generative AI within the NOUS EU project is promising. This expansion not only indicates a commitment to continuous improvement but also presents an opportunity to address any concerns raised in the post-survey. Anticipating an improvement in satisfaction levels with the introduction of new features is a positive outlook for the future of the Transport Cloud concept.

2.6. Type of data used by the respondents

The observation of a similar distribution of answers regarding the type of data respondents deal with in both the pre and post-surveys is intriguing, as it suggests consistency in the profile of data consumers across the two rounds of interviews. This similarity in distributions implies that the respondents' roles and data-related activities remained relatively stable over time.

The similarity in responses indicates a consistency in the profiles of data consumers participating in both survey phases. This stability could be influenced by the nature of the projects, tasks, or roles that participants are engaged in, demonstrating a continued relevance of the Transport Cloud across various data-related contexts.

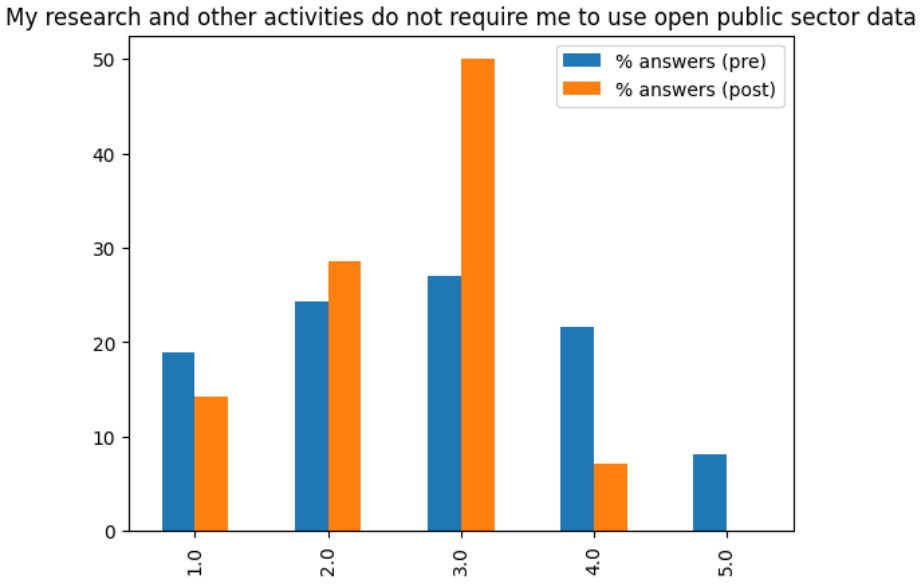


Figure 15: Do not require to use open public data

The acknowledgment of limited use of public open data suggests that respondents' work involves a mix of both open and private data sources. This insight is crucial for understanding the diverse data landscape within which participants operate. The need for a combination of open and private data aligns with the nuanced requirements of real-world data science and research activities.

The heterogeneous distribution of answers in both survey phases highlights the diversity of data types participants need for their research and work activities. This diversity underscores the complex and multifaceted nature of data usage in various domains, necessitating flexible and adaptable data-sharing platforms.

Recognizing the varied dataset types respondents work with provides valuable insights for enhancing the Transport Cloud. Understanding the diversity of data requirements can inform platform developers about the need for flexibility in accommodating different types of data, including private and specialized datasets.

The heterogeneity in responses emphasizes the importance of a user-centric approach in platform development. Adapting the Transport Cloud to cater to the specific and varied data needs of users can significantly contribute to overall user satisfaction.

When considering data accessibility within the Transport Cloud, a notable finding is that in 50% of cases, users can download datasets without the requirement of registering. However, a substantial portion of datasets (30%) necessitates user registration for access. Interestingly, this pattern of dataset accessibility remained consistent between the pre and post-interview phases, indicating stability in this aspect across the project timeline.

The fact that 30% of datasets mandate user registration suggests a deliberate choice in balancing open access with the need for user identification and tracking. Registration requirements can be implemented for various reasons, such as ensuring responsible data use, tracking user engagement, or adhering to privacy and security standards.

The combination of open access and restricted access through registration demonstrates a balancing act between promoting data openness and maintaining control over dataset distribution. This balance is crucial for fostering collaboration while addressing concerns related to data privacy, security, and responsible use.

Funded by the European Union

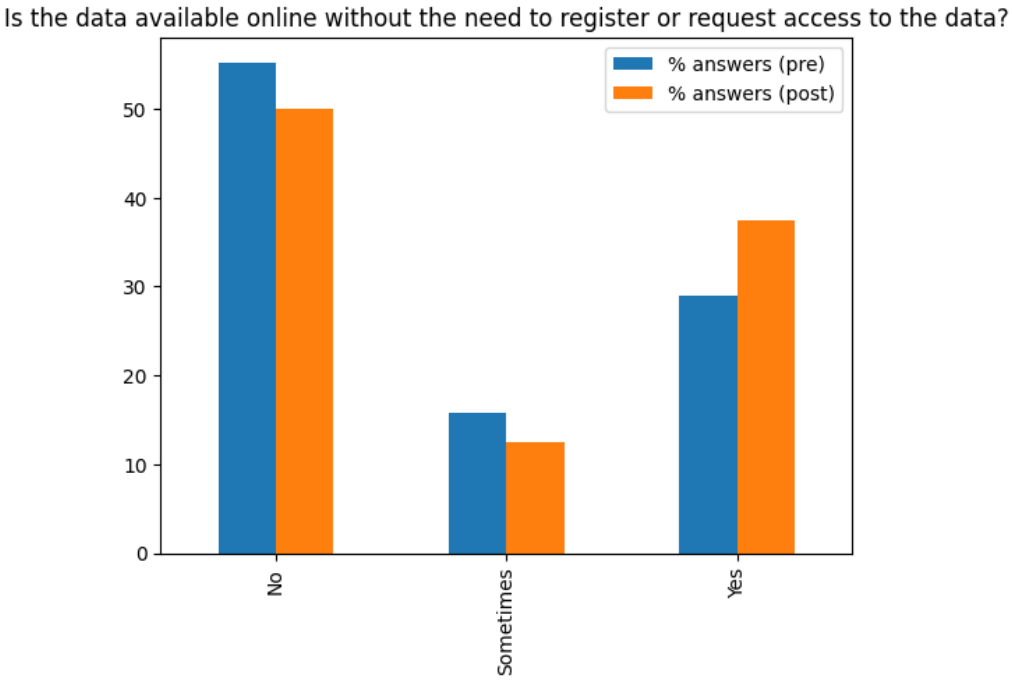


Figure 16: Data available without the need to register

The 50% of respondents indicate that data is available free of charge. This is an interesting finding, particularly when considering the use of the Virtual Labs platform during the MobiDataLab events.

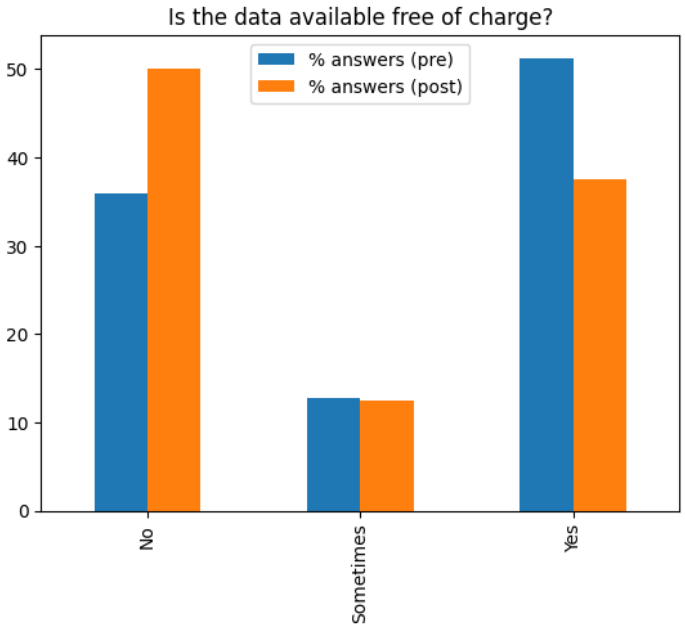


Figure 17: Data available free of charge

The change of 10% in the availability of open data between the pre and post-surveys implies a certain level of dynamism in the data landscape. This could be influenced by evolving datasets, updates, or modifications made to the Virtual Labs platform, demonstrating a responsive approach to user needs. The use of Virtual Labs as a platform during MobiDataLab events, coupled with its provision of an open transport and mobility data catalog, aligns with the respondents' perception of data being available for free. The catalog's role in showcasing open data contributes to users' awareness and utilization of freely accessible datasets.

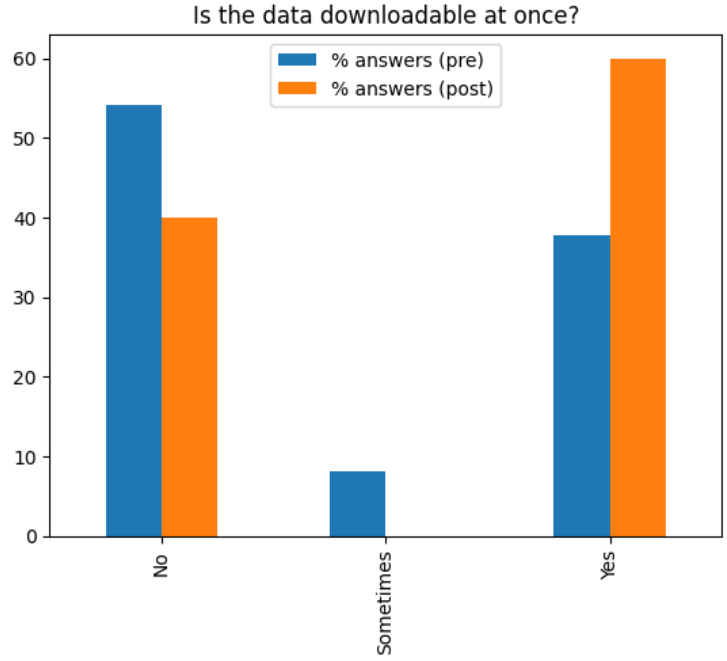


Figure 18: Data downloadable at once

The reported increase from 55% to 60% in respondents claiming the possibility to download datasets at once between the pre and post-surveys is a notable trend. This observation sheds light on the user experience aspect related to the ease and efficiency of downloading datasets, which is a crucial factor in data-driven research and analysis. The increase in the percentage of respondents claiming the ability to download datasets at once suggests potential improvements or optimizations in the download process over the project timeline. This could be a result of platform enhancements, user feedback implementations, or adjustments to data distribution methods. The positive shift in respondents claiming the possibility of downloading datasets at once is an indicator of potential increased user satisfaction. This aspect is crucial for the overall success and adoption of the platform, as a positive user experience is closely linked to user engagement and continued usage.

The last three questions deal with the characteristics of the dataset. Most respondents (more than 60%) agree that data are in machine-readable format. The percentage of answers is very similar between the pre and post-survey. The similarity in the percentage of respondents across both survey phases reflects a stable perception of the dataset characteristic related to machine readability. This stability is crucial, as it indicates that users consistently perceive the data to be in a format that can be easily processed by machines.

The agreement on machine readability aligns with best practices for user-friendly data access. Machine-readable formats enable automated processing and analysis, facilitating seamless integration into various data-driven workflows, tools, and applications.

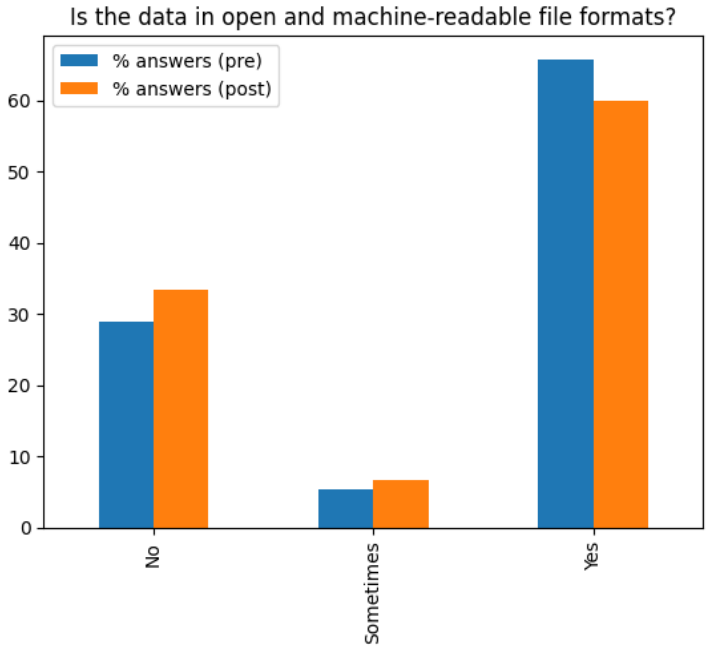


Figure 19: Data in open and machine-readable format

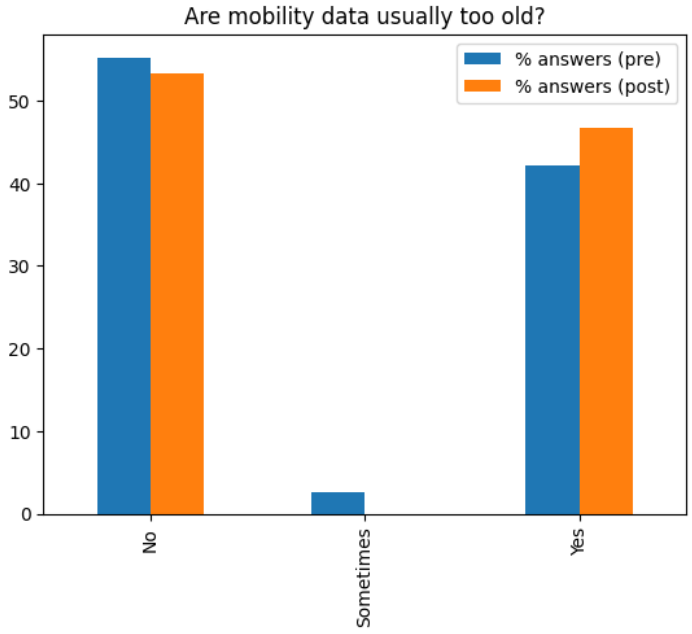


Figure 20: Data usually too old

Machine-readable formats contribute to data interoperability, allowing users to integrate diverse datasets into their analyses. This characteristic is particularly important in collaborative research settings where datasets from different sources may need to be combined.

The even distribution among respondents, with about half believing that the data is too old or must be updated frequently, and the other half disagreeing, suggests a nuanced perspective that aligns with the diverse nature of datasets and data providers. This balanced distribution between the pre and post-interviews indicates a consistent perception among users over time.

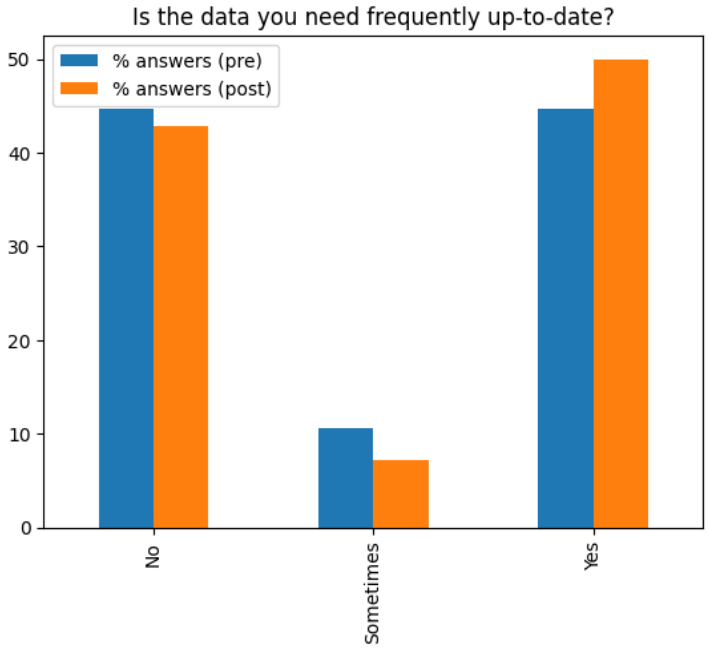


Figure 21: Data needed to be frequently updated

The perception of whether data is too old or requires frequent updates is inherently subjective and can vary based on the specific needs and expectations of users. The even distribution among respondents likely reflects the diverse range of data types and applications within the Transport Cloud.

Different datasets serve various purposes, and their temporal relevance varies.

Some datasets may require frequent updates (e.g., real-time traffic data), while others may have a slower update frequency without compromising their utility.

3. Evaluation of the data exchange culture

3.1. Methodology

We propose a methodology to identify the claims and the main topics in a set of interviews to understand the data-sharing culture in the transport sector.

Figure 22 shows the main components of the proposed methodology:

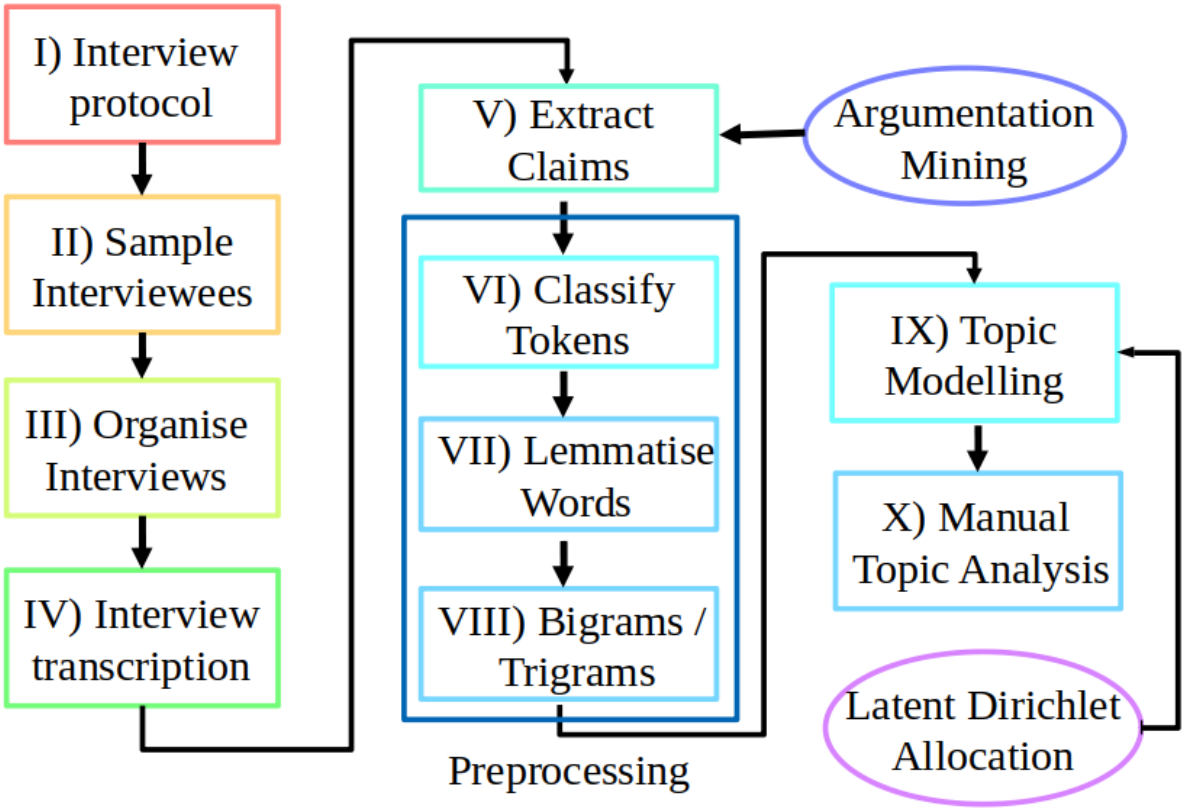


Figure 22: Methodology to analyse interviews on data-sharing culture

The methodology is articulated as follows:

1. Define Interview Protocol (I): Initially, we define a protocol for the interview with macro-categories of questions. These questions are identified thanks to the analysis of the literature and based on the goal of the work.

2. Identify Heterogeneous Sample (II): Next, it is necessary to identify a heterogeneous sample of interviewees based on some macro-categories of actors involved in the transportation sector to ensure a diverse and representative sample by including participants from various roles within the sector.

3. Contact and Schedule Interviews (III): Reach out to the identified interviewees via email to organize one-hour interviews. The scheduling process is a crucial step in ensuring effective participation and gathering diverse perspectives.

4. Conduct, Record, and Transcribe Interviews (III): The interviewee is contacted via email to organize a one-hour interview. We record the interview and transcribe it for textual analysis.

5. Argumentation Mining (V): First, we extract claims and evidence from the input text using the off-the-shelf argument mining tool MARGOT. MARGOT is an online server that extracts argument portions (claims and evidence) from any textual document as input. Although the underlying model uses Wikipedia articles, it has shown its performance across different genres, topics, and domains (Lippi & Torroni, 2016; Lippi et al., 2022). Argument component detection is performed by MARGOT using tree kernels (Moschitti, 2006) that exploit the structure of constituency trees to look for similarities between sentences. The classifier computes two distinct scores for each sentence, which is the confidence assigned by MARGOT to the fact that such sentence contains a claim or, respectively, evidence. By default, if the corresponding score is positive, a sentence includes a claim (respectively, evidence).

6. Preprocessing Text (VI-VIII): Preprocessing tasks include tokenization, part-of-speech tagging, and entity recognition (VI). We use spaCy and Gensim to tokenize words, assign tags to them, and implement the entity recognizer to classify tokens according to the transition-based algorithm (Covington, 2001). We lemmatize words to reduce words to their base or root form (VII) and tune the thresholds to determine the bigrams and trigrams to enhance the identification of meaningful phrases (VIII).

7. Topic Modelling (X): Finally, all the argument portions detected in the first stage and preprocessed are analyzed via topic modeling to identify relevant topics. We created a dictionary of data and a corpus with terms frequencies to implement Latent Dirichlet Allocation (LDA) which is a generative probabilistic model widely used for topic modeling, a natural language processing (NLP) technique that uncovers underlying topics within a collection of text documents. Implement Latent Dirichlet Allocation (LDA) to create a representation of words that allows predicting other words in a phrase or document based on contextual information. Developed by David Blei, Andrew Ng, and Michael Jordan in 2003, LDA is based on the assumption that documents are mixtures of topics and topics are mixtures of words. LDA has proven to be a powerful tool for understanding and extracting meaningful patterns from large textual datasets, contributing to various applications in natural language processing and information retrieval.

Funded by the European Union

Gensim is a popular Python library for natural language processing tasks, and one of its functionalities is phrase detection. It leverages the Skip-gram model, an extension of the word2vec model introduced by Mikolov et al. in 2013.

The Skip-gram model is designed to create word representations that predict other words in a given context. The Skip-gram model is a type of word embedding model that learns distributed representations of words in a continuous vector space. It aims to predict the context words (surrounding words) given a target word in a phrase or document. This model captures semantic relationships and similarities between words based on their co-occurrence patterns. The phrase detector in Gensim is inspired by the work of Bouma in 2009. The connection of words into phrases occurs based on their co-occurrence patterns and the probability that one word appears given the occurrence of another word. The probability equivalence ensures that two words are connected if they tend to appear together and have similar probabilities of occurrence.

The combined use of Argumentation Mining and Topic Modelling provides a robust framework for processing and extracting valuable information from interview transcripts. This approach enables the identification of claims and main topics within the data-sharing culture in the transport sector.

The rationale behind combining argumentation mining and topic modeling is to filter out irrelevant content using argument mining. Focusing on identifying claims and evidence acts as a pre-processing step to retain only the most pertinent information for further analysis. Topic modeling on the argument portions allows for a more focused exploration of relevant topics, enhancing the interpretability of the extracted insights. The combination of argumentation mining and topic modeling leverages the strengths of each technique. Argumentation mining narrows down the focus to argumentative elements, offering a structured way to identify claims and evidence. Topic modeling provides a higher-level analysis, revealing overarching themes and topics within the extracted argumentative content.

This approach is likely to enhance the efficiency of extracting valuable insights from large volumes of text by first identifying argumentative structures, it is possible to prioritize the most relevant content for subsequent topic modeling. The synergy between these NLP techniques can lead to a more refined and focused analysis, providing a clearer understanding of the underlying discourse. In summary, the approach demonstrates a thoughtful integration of argumentation mining and topic modeling, highlighting the synergistic benefits of combining these NLP techniques for a comprehensive analysis of textual data in the context of understanding claims and topics.

We manually performed several experiments and checked the coherence indicator and the frequency of words on each topic. According to the frequency of words occurring in each topic, we removed the ones such as maybe, example, and super, we also considered if the same word appeared in different topics. The experiments consisted of the α parameter manual tuning - which we finally set to 0.001 and the number of topics.

Finally, we manually analyze and discuss the identified topics and report some claims as examples.

Funded by the European Union

3.2. Description of the Dataset

This study utilizes a dataset derived from interviews conducted between June 2022 and September 2023. The interviews engage with key stakeholders who represent the data-sharing culture within the transportation sector. The participants in these interviews encompass a diverse range of actors, including policymakers, representatives of municipalities, transport operators, providers of Intelligent Transport Systems, software providers, and car manufacturers, as outlined in Table 1.

Table 1: Number of interviews per actor

ACTORS' MACRO-CATEGORIES	NUMBER OF INTERVIEWS
RESEARCH CENTERS/ UNIVERSITIES	2
CITIZENS	2
TOURIST AGENTS	1
CLIMATE CHANGE NGO	1
POLICYMAKERS	3
MUNICIPALITIES/ASSOCIATIONS OF MUNICIPALITIES	3
GOVERNMENT/MINISTRIES	1
PUBLIC TRANSPORT AUTHORITY	3
RAIL INFRASTRUCTURE AUTHORITIES	1
TRAFFIC MANAGEMENT CENTRE	2
TRADE ASSOCIATION	1
LOGISTICS OPERATORS	3
PUBLIC TRANSPORT OPERATORS / TRANSPORT AGENCIES	4
AIRLINES	1
INTELLIGENT TRANSPORT SYSTEMS (ITS) PROVIDERS	4
SOFTWARE PROVIDERS	1
SEARCH ENGINES	1
MOBILE PHONE OPERATORS	1
CLOUD PROVIDERS	1
SATELLITE OPERATORS	1
OEM/CAR MANUFACTURERS/ASSOCIATION OF CAR MANUFACTURERS	2
EMERGENCY SERVICES	1

Each category includes the following types of actors:

- Research and Academic Institutions: Universities and research centers conducting studies and research related to transportation, urban planning, and mobility.
- Tourist Agent: Includes travel agencies, tour operators, and organizations involved in promoting and facilitating tourism.
- Climate Change NGOs: Non-governmental organizations focusing on environmental issues, including those related to sustainable transportation and climate change.

Funded by the European Union

- Policy Makers: Entities responsible for formulating and implementing transportation policies and regulations at local, regional, and national levels.
- Public Transport Authorities: Organizations overseeing and managing public transportation services, including buses, trains, subways, and trams.
- Railway Infrastructure Authorities: Entities operating and maintaining rail transport services, including passenger and freight rail.
- Trade Associations: Entities acting as organizations that represent the interests of businesses within a specific sector.
- Traffic Management Centers: Centers responsible for monitoring and managing traffic flow, controlling signals, and addressing congestion issues.
- Logistics Operators Organizations involved in the transportation of goods, managing supply chain logistics, and facilitating freight services.
- Infrastructure Operators: Entities managing and maintaining transportation infrastructure, such as roads, bridges, tunnels, and railways.
- Public Transport Operators: Organizations responsible for providing public transportation services to the general population.
- Aviation Industry: Includes airlines, airports, and air traffic control agencies involved in air transportation.
- Intelligent Transport Systems (ITS) Providers: Companies offering technology solutions to enhance transportation efficiency, including traffic management systems and smart infrastructure.
- Software Providers: Companies developing software, applications, and technologies for transportation management, data analytics, and mobility solutions.
- Search Engines: Platforms and tools designed to help users find relevant information related to transportation, travel, and logistics.
- Mobile Phone Operators also known as mobile network operators (MNOs) or wireless carriers: Companies that provide mobile communication services to consumers and businesses.
- Cloud providers: Companies that offer cloud computing services, allowing individuals, businesses, and organizations to access and use computing resources over the internet.
- Satellite operators: Satellite operators are companies or entities responsible for the design, launch, operation, and maintenance of satellites in Earth's orbit
- Automotive Industry: Original Equipment Manufacturers (OEMs) and companies involved in manufacturing automobiles, including cars, trucks, and buses.

Funded by the European Union

- Maritime and Port Authorities: Organizations overseeing ports, harbors, and maritime activities, including shipping and navigation.
- Emergency Services: Entities such as police, fire services, and ambulances are involved in emergency response and incident management.

The interviewer initiates the discussion by introducing the MobiDataLab project and outlining the associated activities.

To ensure comprehensive data gathering and analysis while preserving participant anonymity, the interview is recorded for subsequent transcription. The interview protocol comprises three distinct sections:

- Section A (Collaboration and data-sharing culture): This section is designed to evaluate collaboration dynamics and the prevailing culture of data-sharing. Questions within this section seek to identify the primary stakeholders with whom the interviewee collaborates in sharing and receiving data. Emphasis is placed on discerning whether these mobility data-sharing stakeholders share common goals and exhibit trustworthiness and benevolence in their data-sharing practices. Additional inquiries explore the significance of past collaborations as a prerequisite for current data-sharing, along with delving into the data-sharing practices within the interviewee's organization. A specific query addresses the perceived distinction between sharing data and sharing knowledge.
- Section B (Data Quality): Focused on data quality, this section aims to characterize transport data conditions and pinpoint challenges associated with sharing high-quality data. The overarching goal is to understand the transformative impact of the evolving data culture in the mobility sector and its potential implications for society and the environment.
- Section C (Social Impacts): This section delves into the social impact of the data-sharing culture. Questions within this segment seek to unravel how data-sharing can instigate social change and elucidate the expectations surrounding such changes. Furthermore, this section explores the attitudes of the interviewee and other actors toward societal and environmental goals. The final question in Section C assesses the impact of external factors, such as the COVID-19 pandemic and the Russia-Ukraine conflict, on integrated mobility solutions, including Mobility as a Service (MaaS).

3.3. Evaluation of the data-sharing culture

In this section, we describe the topics identified thanks to argumentation mining and topic modeling in the context of the overall methodology previously described.

Topic I: Data quality and completeness

The discussions on data quality and completeness within the context of mobility underscore the intricate interplay between accuracy, representativity, and the open availability of data. Data providers should be capable of delivering accurate data.

Funded by the European Union

However, having accurate data is a starting point rather than an endpoint, as there is the need to anonymize them without compromising quality.

Issues surrounding data quality and completeness are intricately woven into the work of data scientists. The understanding of these nuances requires advanced skills in data science to navigate potential pitfalls, ensuring that the dataset is not only complete but also meaningful from a transport perspective. The ongoing work involves grappling with potential errors during data analysis, particularly when anonymization is not executed correctly. This highlights the risk that poorly anonymized data might compromise the quality of the dataset.

To improve the overall quality of data, it is important to have a direct connection with those who own the data. The knowledge of data limitations becomes crucial, particularly when dealing with different providers offering datasets of varying comprehensiveness. Assurances and clear expectations are essential in this work, ensuring that end-users fully understand the capabilities and limitations of the provided data.

The quest for completeness extends beyond accuracy to encompass the regular availability of data. There is the need for complementary solutions to improve the manual collection and analysis of data, rather than a complete substitution. Moreover, data quality is not uniform across all providers, with smaller data providers potentially facing challenges in responsiveness compared to larger counterparts.

Topic II: **Legal responsibility**

The challenges and considerations surrounding data-sharing in the realm of public services are ever-present, as ensuring legal compliance and clarity remains paramount. The legal responsibility associated with public datasets demands a clear understanding to prevent potential risks, especially considering the evolving landscape with regulations like GDPR. It is always a learning process, adapting to changes in data protection mechanisms and navigating the delicate balance between fostering data-sharing and safeguarding individual rights.

The complexity becomes apparent when datasets are sufficiently large, requiring the application of machine learning algorithms. The need to work with data to extract knowledge highlights the constant learning curve in the domain of data-sharing. Clear communication is key, especially when engaging with partners from the public sector, private sector, and universities. Emphasizing the importance of collaboration with citizens is vital to ensure correct and legal data-sharing practices.

Data-sharing in the context of mobility reveals two somewhat independent ecosystems—one centered around personal vehicle data and the other around public transport data. The goal is to strengthen data-sharing based on an open data approach, particularly where public information about transport alternatives or delays is concerned. Simultaneously, efforts to enhance data protection and sovereignty for individuals participating in the mobility system underscore a commitment to transparency and personal data control.

In the collaboration between the public and private sectors, differences in mindset emerge, with the private sector leaning towards commercialization and revenue generation. However, both sectors share a common goal of ensuring an interconnected and transparent transport system that serves people seamlessly across regions.

To tell people about the economic and state authorities' use of their data is crucial to strengthen public knowledge about data-sharing. This transparency serves as a foundational element in building trust and fostering a clear understanding of the role of data-sharing in public services. Ultimately, creating a place where people feel empowered to make decisions about their data contributes to the overarching goal of an inclusive, interconnected, and transparent transport system that benefits everyone.

Topic III: **Collaboration and trust**

Establishing trust is a key element in the collaborative landscape, particularly when dealing with diverse stakeholders and end users. End-user engagement plays a crucial role in fostering this trust, though, there are frustrations at times.

When it comes to collaboration with stakeholders, understanding the limitations of trust is paramount. Governmental authorities serve as a trusted third party, covering a pivotal role in ensuring trust among all stakeholders. They play a key role in building and maintaining trust between various data holders, facilitating a collaborative environment where everyone is assured of the integrity and reliability of shared data.

However, trust can be strained, as evidenced by instances where collaboration falls short of expectations. In cases where partners do not respect the research question and provide data of quality, it becomes evident that future collaborations with such entities may not be fruitful. This underscores the importance of trust as a key factor in collaborative endeavors.

Choosing the right partner is a nuanced process, especially when considering various projects and activities. While the desire is to collaborate with all partners, practical limitations necessitate selecting those with the best ideas, a timely approach, and a profile suitable for the specific job at hand. The question of partner selection, therefore, becomes a delicate balance of fairness, encouraging active participation, and diversifying opportunities, even for those who may not have been highly active in previous collaborations.

In essence, trust is a dynamic element in the collaborative landscape, particularly as we navigate the challenges of demographic representativeness, frustrations, and the selection of partners. Recognizing the key role of trust, both in engaging end users and collaborating with stakeholders, remains fundamental to fostering successful and sustainable partnerships in our initiatives.

Topic IV: **Previous experience and knowledge-sharing**

In the realm of data-sharing, creating value is not just a short-term gain but a long-term commitment. It is essential to better communicate the positive aspects of sharing data, drawing inspiration from previous experiences and case studies. Companies, particularly, find value in learning from past successes and understanding how sharing solutions or data sources openly has contributed to meaningful outcomes.

Bringing transparency back to the table is crucial for building trust and fostering collaboration. It is not just about showcasing the successes but also being honest about the lessons learned. The important aspect is not just looking within our own teams but across the organization to identify individuals with the best knowledge and experience to contribute to a data-sharing project.

This approach is vital for ensuring that knowledge is shared openly, creating a collaborative environment where everyone's expertise contributes to the project's success.

The reality of the process evolving and mindsets changing is an important factor to consider. Experience and learning from the past are valuable assets, shaping how we adjust the scope of our projects to align with these evolving insights. This adaptability is crucial for the long-term success of data-sharing initiatives.

Participating in EU projects adds another layer of motivation for data-sharing. Beyond the funding received, the experience gained from collaborating with stakeholders from different countries, traveling, and visiting various places creates a sense of commitment.

This commitment stems from the human aspect – the satisfaction of working on projects that bring value not just in the short term but contribute to the objectives of the project in the long term.

Topic V: **Social and environmental impact**

In the pursuit of sustainable transportation, optimizing logistics and delivery is a crucial way to minimize the environmental impact. This approach not only benefits the environment but also has a significant social impact, contributing to the increased satisfaction of citizens. By providing citizens with diverse choices and reducing commuting times, it aligns with the expectation that citizens can optimize their travel, thus fostering environmental awareness.

Publicizing the benefits of using public transport, particularly through datasets and tools that promote various transportation modes, improves congestion. As the cost of using personal cars, notably due to high gas prices, becomes a limiting factor, promoting efficient and cost-effective public transport options becomes essential for individuals and organizations alike.

Data-sharing is seen as a catalyst for societal change, particularly in demonstrating the efficiency and environmental benefits of alternative transportation modes over personal cars. To achieve this, demonstrating time and money savings, along with environmental benefits, is deemed a challenging but crucial aspect of promoting public transport.

Interoperability of data is identified as the first step toward achieving this goal. Combining transport data with environmental data enables robust analytics and learning, facilitating a comprehensive assessment of transportation emissions. However, stakeholders acknowledge that achieving environmental goals requires a fundamental shift away from the car-centric approach.

The emerging mobility-as-a-service solutions are considered a theoretical but promising way to use various modes of transport more efficiently. The practical impact on the environment, however, depends on end users' criteria and choices, introducing complexity into the equation.

Machine learning emerges as a powerful tool to ensure proper data usage and recommend environmentally friendly modes of transport. This approach can play a significant role in controlling environmental factors associated with transportation choices.

In any effort to reduce the impact of transport on the environment, rigorous measurement of that impact through data analytics is deemed integral.

Funded by the European Union

Such initiatives require collaboration between companies, organizations, and regions to create a unified approach toward sustainable transportation. The success of such projects depends on different stakeholders working together, aligning expectations, and collectively contributing to a positive environmental impact over time.

Topic VI: **Motivations of data-sharing**

Sharing information with diverse communities of stakeholders is essential for achieving common goals, particularly in the context of building necessary infrastructure. However, when considering aggregating data to uncover trends or clusters of information and employing state-of-the-art AI algorithms, challenges arise.

Effective communication and information exchange play a pivotal role in promoting the positive aspects of data-sharing. In the mobility sector, particularly in public transportation, the challenge lies in attracting individuals with expertise in artificial analytics to work within public institutions rather than opting for roles in the private industry. Bridging this gap is essential for leveraging analytical capabilities to enhance public mobility planning.

Research patterns and modeling, utilizing transportation data, offer valuable insights. However, advancements in planning through data analytics are essential for optimizing resources.

In the realm of data-sharing, considering the diverse needs of stakeholders is paramount. Recommendations for improvement often relate to how information is provided and utilized, especially in the context of environmental considerations. The ability to recommend sustainable practices and policies requires robust data analytics, which, in turn, relies on collaborative efforts and effective communication among stakeholders.

Ultimately, sharing information, fostering communication, and leveraging analytics are integral components in realizing common goals and building the necessary infrastructure. These endeavors contribute not only to the efficiency of transportation systems but also to the broader objectives of environmental sustainability and community well-being.

Table 2: Topics and extract of the interviews

TOPIC NAME	CITATION
TOPIC I: DATA QUALITY	'[data] have to be the most complete as possible ..',
	'[data] has to be accurate. Also, it is important for data the way they reflect and represent people moving around. Completeness is very important and representativity.'
	'So if you anonymize the data in a wrong way, there is a high risk that the results in the dataset are not good quality and usually, I come up with errors when analyzing the data.'
	"Everybody we deal with is completely reliable and find some of the smaller data providers might not be able to be quite as responsive in terms of fixing any issues as the large providers".
TOPIC II: LEGAL RESPONSIBILITY	'..prevent services from going further into data-sharing because we are not sure if what we are doing is legal..'
	'..we need to ensure that errors within datasets, which are made public by public administrations will not result in legal responsibilities of the public sector.'

Funded by the European Union

	“Things are changing with the GDPR and any kind of data protection mechanisms”.
	“When we discuss with our partners, public authority, private sector, and university working together we need you to work together with the citizens to make sure that the data-sharing is correct.’,
	'..there are two ecosystems that work pretty independently, one centered around car data or personal vehicle data, and the other centered around public transport data.',
	'We want to strengthen and enhance data-sharing based on an open data approach when we think that data should belong to the public, like information about current transport alternatives or delays in the traffic, but we also want to strengthen data protection when individuals data is concerned and data sovereignty for the people taking part in the mobility system so that they can decide on their own what happens to their data [about] personal activities .',
TOPIC III: COLLABORATION AND TRUST	"We do have a team that looks after the engagement with the end users and I am sure that there are sometimes some frustrations, we might not get as many responses as we would like."
	"you start to collaborate with [stakeholders], trust them by knowing also the limitations of trust. [Governmental authorities] are a trusted third party, so everybody should trust them, and they must trust everybody as they need to ensure that all stakeholders work together. [Governmental authorities] are building trust between all kinds of data holders so that they start working together ."
	"..it is obvious we will not collaborate with this partner again because they do not respect the research question and the data sample they have collected is not of very high quality"
	"If the question is more how to choose a partner, that is trickier because we want to collaborate with all our partners, but some projects and activities you can only have a certain number and can be the ones with the best ideas, the one that comes earliest, the ones we think a certain profile is needed for a certain job and have to try to be fair to our partners, even those who have not been too active, we try to invite them in and not always have the same ones ."
TOPIC IV: PREVIOUS EXPERIENCE AND KNOWLEDGE-SHARING	"..better communicate what would be the positive aspects of sharing the data and it is always, at least for the companies, the best case is from previous experiences, case studies, how things have been done in the past and that can attract them to join, for instance, providing their solutions or data sources openly to others.",
	“So bringing back to the table it is very important and also on the actual honesty and being transparent side of things is equally important so, in terms of the way we work, we have got a vast amount of knowledge experience … and rather than looking at our team, we will look at across the organization, who has the best individual to be working on that project because they will bring their knowledge in with us and it would be good to think that other organizations think the same way as well, bringing their knowledge onto a project and being open and transparent in terms of what's good, what has not… I think the lesson learned side is very important as well in terms of projects .., so bringing in all that knowledge is quite important in terms of the actual success of our data-sharing project.",
	"Reality is that the process evolves, mindsets evolve, which is a good thing because we evolve from experience, we evolve from learning, learnings from the previous years, but that is also the reality that you have to adjust the scope to your learnings and .",

Funded by the European Union

	"I think the EU project gives very good motivation for data-sharing because all organizations participating are motivated not only by the funding received, but also from the experience they have in the EU projects which implies meeting stakeholders from other countries, traveling countries, visiting places, etc. we are humans, and If they are happy to work they are more committed to reaching the objectives of a project .",
TOPIC V: SOCIAL AND ENVIRONMENT IMPACT	'..you optimize logistics and delivery to make sure that their impact on the environment is as small as possible..'
	'The social impact is the increased satisfaction of the citizens that is important for the city. For instance, they do not have to spend too much time commuting, they can optimize, they have different choices, and they are also environmentally aware of the impact of their choices.',
	".. if more people are using public transport and if public transport is better for people. Employment, announcing public transport obviously will also enhance employment by allowing people to be able to go to work without having to take the car, at the moment, using your car is very expensive because of the price of the gas. If using public transport datasets, the right tools to promote public transport and to make them available to people so they can see there is a way to go from their home to work using public transport, even, other types of transport, like on-demand transport like car sharing, ride sharing and so on.",
	'I think that [data-sharing] can change the society."
	'For social impact, I think that you have to demonstrate that you live your transport, you have to demonstrate that Public Transport is more efficient, more time-saving, more environmentally saving than your car, you have to demonstrate time-saving, money-saving, and environmental saving I think is the most difficult thing to do.'
	"So the first thing is that we have to make the data interoperable so that you can combine them and once you can combine transport data with environmental data, then you can have good analytics and good learning possible to assess the emission of the transportation.",
	"They think we can achieve environmental goals, but not necessarily in the long term. I mean, the only way that we are going to achieve that is by getting this shift away from the car.",
	'So many items being shipped all together, so you optimize logistics and delivery to make sure that their impact on the environment is as small as possible .',
	mobility as a service is supposed to be the next big thing, so again, theoretically, you will be able to use more modes of transport being integrated and you will be able to choose the way you want to travel, more efficiently, in terms of impact on the environment. Well, I guess it depends on what critic the criteria that end users use practically to choose their modes of transport, so do they choose the quickest way? So I do not think there is one answer that mobility as a service could have a positive impact on the environment because it may be the case that everyone chooses the cheapest version, but the cheapest version might still have a big impact on emissions and the environment.",
	'We can use that machine learning to make sure we are using that data properly and recommending different modes of transport to users will help control those environmental factors .'
	'Any effort to reduce the impact of transport on the environment has to start with a rigorous measurement of that impact, so data analytics is integral.
TOPIC VI: SUCCESS OF DATA-SHARING	'But if you think about aggregating the data to build trends or clusters of information and you want to utilize state-of-the-art AI algorithms for instance and if you cannot provide the low-level details to the data processing, you will not be able to find out the best results because you have to, for instance, aggregate or anonymize the data if there are people involved.',

'Sharing information with different communities of stakeholders and working with them to build the infrastructure they need is useful for common goals .'
I think it is mostly down to the communication and inform exchange, to better communicate what would be the positive aspects of sharing the data.”
'I think the mobility sector, especially the public mobility sector, has the challenge to get people with knowledge in artificial analytics and so on that work at a municipality, at a state-run institution or something and not go to the private industry.',
"Of course, there are things like research patterns that can be done which is using maybe transport things and you can do it now, that is a model, but of course, there are a lot of improvements in our planning with data analytics, we have fewer trains, which runs empty on our network to be on the right place at the right time."

Funded by the European Union

4. Conclusions

In conclusion, the assessment of the Transport Cloud's acceptance reflects a predominantly positive sentiment among survey participants. The overall satisfaction underscores the platform's success in fostering a collaborative environment and enhancing information exchange within the transportation sector. However, while the Transport Cloud has made substantial strides, there are specific aspects that may not have met the heightened expectations of some participants.

This underscores the importance of managing expectations and ensuring that the Transport Cloud's capabilities align closely with the envisioned benefits. Addressing these concerns can contribute to a more comprehensive and satisfactory user experience.

This assessment can serve as a foundation for continuous improvement, allowing for refinement and optimization of the Transport Cloud. By addressing potential gaps or unmet expectations, the platform can evolve to better cater to the diverse needs of its user base, ultimately solidifying its position as a valuable tool in fostering a collaborative and efficient data-sharing culture within the realm of transportation.

Finally, in this work, we identified the main topics of the data-sharing culture that emerged from the interviews.

We found that ensuring data accuracy, completeness, and representativity is critical. However, anonymizing data incorrectly can jeopardize its quality. Addressing challenges with smaller data providers and emphasizing the importance of accuracy and responsiveness are key to maintaining high data quality standards.

Furthermore, legal uncertainties can hinder data-sharing initiatives. Mitigating legal risks and responsibilities, particularly in the context of GDPR and data protection mechanisms, is essential. Collaboration with governmental authorities and the need for trust in data-sharing underscore the complex legal landscape.

Effective collaboration with stakeholders, including governmental authorities, is crucial. Building and maintaining trust among all data holders are fundamental for successful collaboration. Partner selection based on project requirements, fairness, and a willingness to learn from past experiences contribute to robust collaborations.

Drawing on previous experiences and case studies is instrumental in communicating the benefits of data-sharing. Transparency, honesty, and knowledge-sharing across organizations foster a culture of openness. Adapting to evolving processes and mindsets ensures the success of data-sharing projects.

Data-sharing can significantly impact society and the environment. Optimizing logistics, promoting public transport, and demonstrating the positive societal and environmental effects are crucial. Interoperability and data analytics play key roles in achieving environmental goals and driving a shift away from reliance on individual vehicles.

Funded by the European Union

Sharing information with diverse stakeholders' aids in building necessary infrastructure and achieving common goals. While emphasizing the positive aspects of data-sharing, effective communication, and collaboration are highlighted as challenges. Encouraging individuals with expertise in artificial analytics to contribute to public mobility initiatives is crucial for success.

In conclusion, the themes covered highlight the multifaceted nature of data-sharing. Legal considerations, collaborative efforts, past experiences, and societal and environmental impacts collectively underscore the importance of responsible and strategic data-sharing practices. The success of these initiatives relies on addressing challenges, fostering collaboration, and leveraging collective knowledge for positive societal and environmental outcomes.

Funded by the European Union

5. References

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

Bouma, G.: Normalized (pointwise) mutual information in collocation extraction. Proceedings of the Biennial GSCL Conference 2009 (01 2009)

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

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

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

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

Funded by the European Union

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