1st Conference on Research Data Infrastructure

Linking RDM

https://doi.org/10.52825/CoRDI.v1i.370

Published: 07 Sept. 2023

Designing a Mobility Data Trustee (MDT)

Findings from a Multi-Disciplinary Analysis of Requirements of an MDT

Andreas Czech[1][https://orcid.org/0000-0002-6895-0606], Vivien Geenen[1][https://orcid.org/0009-0009-7899-8249], Constantin Breß[2][https://orcid.org/0000-0002-2133-1541], Marija Turkovic Popovski[2][https://orcid.org/0009-0007-0400-151X], Peter Krauß[1][https://orcid.org/0000-0002-5869-352X], Till Riedel[1][https://orcid.org/0000-0003-4547-1984], Frank Gauterin[1][https://orcid.org/0000-0002-0870-7540]

[1] Karlsruhe Institute of Technology, Germany

[2] FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur, Germany

Abstract. A large amount of different data is currently collected in the mobility sector. However, due to technical, legal, and economic hurdles, it cannot be made usable. The "TreuMoDa" project, in which a Mobility Data Trustee (MDT) is being designed and tested as a prototype in the autonomous driving test area in Baden-Württemberg, provides a solution to this problem. Data trustees are a pioneering option to enable cross-sectoral data exchange between different actors from industry, science, and society. We provide an insight into the concept of the MDT. Specifically, the requirements of various stakeholders from the automotive industry, software developers, infrastructure, cities, and transport companies are analyzed with regard to the expected organizational, legal, and technical functions of such an MDT. The concept of a data trustee is particularly relevant to research data infrastructures, as it can enable the flow of data from research to industry and vice versa. Our work therefore also will look at the MDT as part of a larger research data management scheme.

Keywords: Data Trustee, Mobility Data, Anonymization, Personal Data, Intermediary, Data Sharing

1. Why the MDT is needed

Data intermediaries are foreseen as enablers for data sharing by recently introduced regulations [1]. However, it is not obvious how trusted third parties can facilitate such data reuse. When looking at the application area of autonomous driving, one challenge is how to deal with personal information contained in necessary data to develop and validate core functionalities of the systems: If data collection and processing under the GDPR is already a challenge, data reuse, and sharing turns out to be an even larger challenge.

1.1 Motivational Example

Consider a simple research project that requires data exchange between two consortium partners: One partner generates camera data from a test vehicle. The other Partner needs the data set to train an object detection model for a driver assistance system that applies brakes automatically when pedestrians cross the road. The system recognizes pedestrian intention. In this case, anonymization of the data is required before data exchange because the data contains the visible faces of people. However, as we can assume, the data provider is neither an expert in anonymization (black boxes would draw the dataset useless) nor in the legal requirements for data exchange (we need to balance the risks of data subjects). The partners

would already require help from an independent body, we call such an intermediary third party that mediates between both sides a Mobility Data Trust (MDT).

1.2 Research Questions

Within our work, we address the following research questions:

(R1) What are the necessary functions an MDT needs to provide?

(R2) Which regulatory, technical, and organizational requirements need to be met by MDTs?

(R3) Are MDTs necessary when purely looking at research data infrastructures?

Based on research questions R1 and R2, we have followed an interdisciplinary requirement engineering approach to design a feasible concept of a data intermediary called the Mobility Data Trustee. Based on our learnings, we will further discuss R3 as an outlook to a future harmonized approach that fits both industrial data spaces and research data infrastructures.

2. Requirements of the MDT

We analyzed both risks and benefits from the organizational, legal, and technical perspectives based on domain-specific constraints and best practices.

2.1 Organizational

In particular, the customer requirements are distinguished depending on the data provider, the data user, or both. Both groups are interested in simple access to data, need trust in the MDT, and education and advice by the MDT in terms of law, anonymization, and the technical part of data sharing (upload, download). Especially, the last-mentioned point needs to be realized through personal contact with the MDT. Data users require applicable data for their points of interest. On the other hand, data providers expect a high level of safety for their own data. Therefore, a strong degree of anonymization is a condition for data sharing. Furthermore, data providers require insistence on control of the choice of potential data users. The data trustee is to function as a not-for-profit data mediator between those sometimes conflicting interests. To achieve a critical mass of users, its service needs to be advertised to actors from different sectors, and communication between them needs to be uncomplicated and transparent.

2.3 Regulatory

The legal requirements of data trustees as intermediaries are governed by legislation on data protection, data governance, and related fields with the main goal to facilitate data sharing. In this regard, the most relevant are the GDPR and DGA, which shape the concept of a data trustee, as well as determine this role and legal responsibilities. The legal analysis deals with the legal review of mobility data and the definition and classification of the anonymization process. This follows the determination of levels, methods, and guarantees of anonymity. Eventually, as data sharing is one of the main drivers of the whole process, the research also addresses conditions for access and transfer of data under the GDPR and DGA.

2.4 Technical

The infrastructure must be able to handle large amounts of data, be interoperable with a variety of common data formats and exchange standards, secure the data at transit, at rest, and in use, and allow for strong data governance policies to ensure secure and GDPR-compliant data management. These challenges are combined to a scalable and operable data platform.

3. Preliminary Concept of the MDT

The most important process is the exchange of personal data from a data provider with a data user as shown in Figure 1. To transfer data of data providers to the data users of their choice, it is needed to previously acquire the necessary consent and make parties familiar with the conditions of the data transfer and usage. An anonymous sample is created to allow the data user to check whether the data is useful for their application. Thereby, the process adheres to the legal requirements for lawful, transparent, and secure processing of personal data, particularly regarding the requirements under GDPR and DGA on the consent of the data subjects, facilitation of data sharing, and responsibility for ensuring legal compliance.

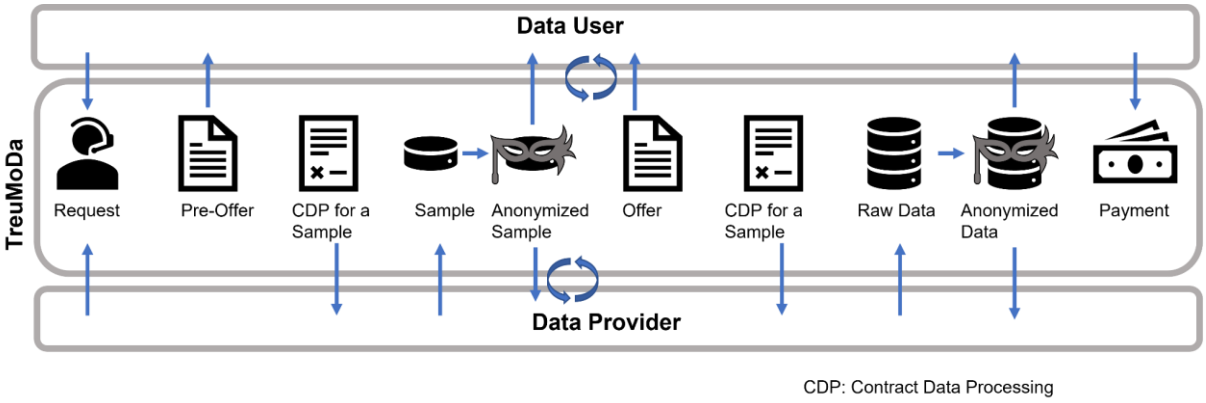

Figure 1. A simplified process of the MDT with the service of anonymization

Other developed processes are supporting data users in finding suitable data providers or checking the anonymization of the data for the current state of the art.

4. Outlook and Discussion

We are currently in the process of evaluating the conceptualized Mobility Data Trustee in the project "TreuMoDa -Treuhandstelle für Mobilitätsdaten". We believe that the concept is of interest to a national research data infrastructure as it may enable the cross-sectoral flow of data from research to industry and vice versa. Making (research) data available anonymously while retaining important aspects of the data is essential for compliance with data protection regulations and effective data reuse particularly between industry and research. The mobility domain has special requirements and faces difficult challenges regarding the handling of personal data that might not apply to all research data. Our initial analysis has shown, however, that the developed technical, legal, and organizational concepts as well as the business model and financing options are transferable. From our research, we have learned that the functions of a data trustee need to be aligned with governance and organizational structures to match regulatory requirements. Therefore, we believe it is important to consider such intermediaries early in the design of a national research infrastructure.

Data availability statement

The submission is not based on data.

Underlying and related material

There is no other material that supports our findings or is closely related to the article/contribution.

Author contributions

Andreas Czech: Methodology, Writing - Original Draft, Writing - Review & Editing **Vivien Geenen:** Visualization, Project administration, Methodology, Writing - Original Draft, Writing - Review & Editing **Constantin Breß:** Conceptualization, Investigation, Writing - Original Draft, Writing - Review & Editing **Marija Turkovic Popovski:** Investigation, Writing - Original Draft, Writing - Review & Editing **Peter Krauß:** Writing - Original Draft **Till Riedel:** Conceptualization, Methodology, Writing - Original Draft **Frank Gauterin:** Funding acquisition, Conceptualization, Supervision

Competing interests

The authors declare that they have no competing interests.

Funding

Funding for the Project “TreuMoDa - Treuhandstelle für Mobilitätsdaten” was provided by the German Federal Ministry of Education and Research (BMBF). You can find more information on the following page, among others: https://www.bildung-forschung.digital/digitalezukunft/de/technologie/daten/datentreuhandmodelle_pilotvorhaben/datentreuhandmodelle_pilotvorhaben.html

Acknowledgement

Special thanks go to Hannah Bürkle from FIZ Karlsruhe, who did various preliminary work within the framework of the responsible work package and beyond. Many thanks also to Ann-Kathrin Dreher for her work in the legal work package.

References

1. Regulation (EU) 2022/868 (Data Governance Act) [2022] OJ L 152.