Paper ID 280

Data Space Mesh: Interoperability of Mobility Data Spaces

Holger Drees[1*], Sebastian Pretzsch[2], Birgit Heinke[3], Dandan Wang[4], Christoph Schlueter Langdon[4,5]

1. Federal Highway Research Institute BASt, Germany

2. Fraunhofer Institute for Transportation and Infrastructure Systems IVI, Germany

3. acatech National Academy of Science and Engineering, Germany

4. T-Systems International GmbH, Germany

5. Drucker Customer Lab, Peter Drucker School of Management, Claremont Graduate University, USA

* drees@bast.de

Abstract

Data spaces as ecosystems for data exchange are emerging as a prerequisite for the mobility sector, particularly as climate change requires new, seamlessly connected and integrated mobility services. Data exchange within a data space happens either on a decentralized basis between two entities or via data platforms as intermediaries. Despite its decentralized nature, today, a data space features a few centralized elements to ensure trustworthiness (e.g., identity provider) and to enhance utility (e.g., vocabulary provider). Therefore, some governance is required in order to facilitate interoperability and data sharing between data spaces. The recently implemented Mobility Data Space (MDS) initiated in Germany tackles this challenge. The paper reports on first results from linking MDS with the German National Access Point for mobility data as well as with the Telekom Data Intelligence Hub focusing on metadata, trust and legal frameworks.

Keywords: Data markets, structures, privacy and ownership; Mobility Data Space

1. Introduction

The exchange of data promises to be part of solutions to so many current and future challenges. The European Commission has recognized the fact that "making more data available and improving the way in which data is used is essential for tackling societal, climate and environment-related challenges, contributing to healthier, more prosperous and more sustainable societies" in the European Strategy for Data and has drawn a target image of a "single European data space – a genuine single market for data, open to data from across the world" [1]. According to the position paper on design principles for data spaces written by the most important data space organizations and projects, "A data space is defined as a decentralised infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles" [2]. With the anticipation of several sectoral data spaces – one of them being a Mobility Data Space –, they are already hinting at the fact that there will not be a one-size-fits-all solution for every conceivable data sector. Instead, the number of data spaces will increase. Also, data platforms are seen as building blocks, which can be part of a data space. Data space concepts should ensure that the growing number of data spaces will not create new silos that hinder data exchange. Key will be interoperability between and across different data space constellations or mesh data spaces.

A data space mesh (or mesh data spaces) expands on the notion of a data mesh [3, 4] beyond an enterprise setting to include novel distributed, federated and decentralized data systems encompassing enterprise boundaries, like data spaces (see "7 data mesh use case examples" [5]). Merriam-Webster defines a mesh as "a web like pattern" or "an arrangement of interlocking links", which corresponds particularly well with the intertwined dimensions of interoperability of systems including technical, semantic, organizational and legal [6]. This paper goes beyond aforementioned theoretical concepts and design principles to describe first endeavours of actually implementing data space concepts within the Mobility Data Space started in Germany and making it interoperable with two data platforms – the German National Access Point for Mobility Data and Deutsche Telekom's Data Intelligence Hub.

1.a. Interoperability aspects for data spaces

Data space concepts allow a single participant to operate in multiple data spaces. This means that data spaces can be nested under certain conditions. These conditions include that participants use "the same minimal set of functional, legal, technical and operational agreements and standards" [2]. While the position paper on Design Principles for Data Spaces [2] uses the term interoperability mainly for technical agreements like data models, formats, APIs, etc., we use interoperability in this chapter for all kinds of requirements to be met before exchanging data between participants of two data spaces or participants of one data space and an existing data platform.

According to the FAIR principles for data management, data should be findable, accessible, interoperable, and reusable (cf. Figure 1) [7]. Although the I specifically stands for interoperability – here mainly for the structure and representation of (meta)data –, all FAIR principles contribute to interoperability in the wider sense which we are referring to.

Interoperability can have various stages. The simplest stage is similar to the first aspect of the FAIR principles: Making data resources and data services findable throughout different data spaces or data platforms requires a common understanding of metadata. A more advanced stage of interoperability throughout data spaces requires a common trust model. It allows for data sharing between participants that originally have little trust in each other. These stages as well as experiences and approaches how to implement them will be presented after setting the scene which data spaces and data platforms are considered.

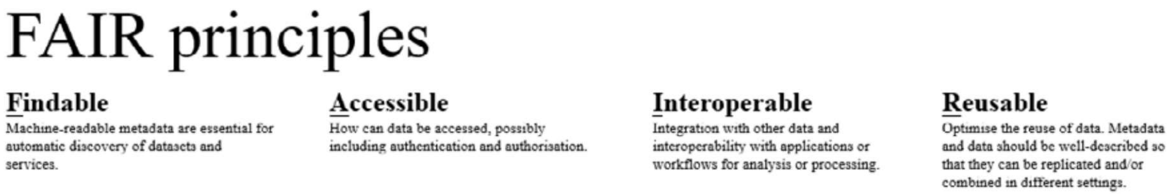

Figure 1 - FAIR principles for scientific data management and stewardship [7]

2. Data spaces and data platforms

In the past, numerous data platforms have emerged, each typically with its own functional scope, its own technical specifications and its own terms and conditions. Efforts have been made to align and harmonize single platforms in order to break up the resulting data silos. In contrast to data platforms, data spaces do not aim for one centralized system which stores and manages all data. Data spaces rather connect different systems which store data which means that they can also connect existing data platforms and integrate them into the data space.

2.a. Mobility Data Space

In 2020, the German Federal Government started to promote and implement the operation of a federated Mobility Data Space (MDS) [8]. A large-scale stakeholder and governance process, led by acatech (German Academy of Science and Engineering) [9], has resulted in an extensive stakeholder engagement, supporting MDS with the provision of mobility data and the implementation of MDS-based use cases. The MDS addresses the private and public sector equally to establish and promote a comprehensive mobility data ecosystem. A very important role will be played by existing data platforms (such as MDM [10], HERE [11], DIH [12]), since they provide access to already connected participants and their data offers. On behalf of the Federal Ministry of Digital and Transport (BMDV), acatech founded the non-profit organization „DRM Datenraum Mobilität GmbH“ in 2021, together with further supporting public and private shareholders [13]. This entity is responsible for the operations of the MDS. On a conceptual and technical level, MDS will provide the central services that are necessary for the operation of a data space according to the International Data Spaces Association (IDSA) [14]: a data marketplace (technically, a metadata directory), a vocabulary provider, an identity provider, a data app store and a clearing house). The data exchange is established directly between participants themselves in a distributed manner by using IDSA-compliant Connectors. The MDS operator has no touching point with the exchanged data itself, resulting in an architecture opposite to data platform/data lake.

The technical concepts to create a Mobility Data Space and progress made, have been previously presented at the ITS World Congresses 2018 [15] and 2021 [16]. The MDS is based on Open-Source reference implementations by Fraunhofer from previous R&D projects on Mobility Data Spaces [17], following the IDSA specifications [14]. Within the MDS, data platforms and data spaces can be made interoperable with IDSA-compliant Connectors. Figure 2 depicts a data space mesh, specifically the interlinking and locking of the two platforms of Mobilithek and Telekom Data Intelligence Hub (DIH) into the MDS via Connectors. Drawing a Connector into a diagram such as in Figure 2 is an easy exercise. In reality, many different and multi-layered interoperability challenges must be overcome. This includes technological, organizational, and legal aspects. So, interoperability will not be binary, like an on-off switch. Instead, there will be various stages or a spectrum of interoperability, for example a light version with interoperability of metadata so that data from one data platform or a data space is visible and can be made available to another data space. And even if two data platforms use the exact same technology, actual data exchange depends on correspondence of legal terms and conditions. As such, interoperability is not given, falling from the sky, but requires governance and enforcement mechanisms. In the metadata example, this could be version control and management of the metadata catalogue.

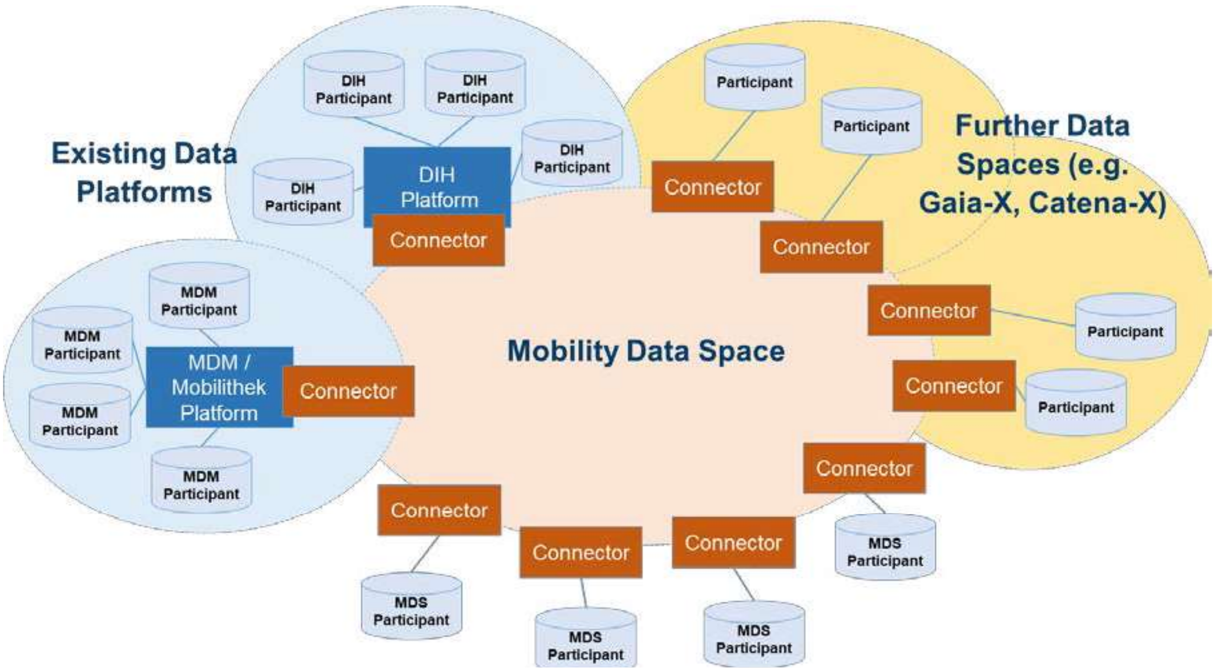

Figure 2 – Data space mesh connecting data platforms and data spaces

2.b. German National Access Point

„Mobilithek" will soon replace the Mobility Data Marketplace (MDM) as the German National Access Point for mobility data (NAP). It is being developed with the clear objective to integrate Mobilithek into MDS. While Mobilithek will provide mainly mobility data from public stakeholders, it is expected that many private entities will be participants of MDS. The intention is to build a bridge and foster data exchange between private and public stakeholders. The same interoperability mechanisms could potentially be used to "connect" all European NAPs in order to commonly contribute to the European Mobility Data Space.

2.c. Data Intelligence Hub

The Deutsche Telekom Data Intelligence Hub (DIH) provides a data sovereignty solution to securely and efficiently exchange, process and analyze data as a platform-as-a-service offering in the cloud. It is composed of (a) Connector, (b) Data Exchange or Marketplace, and (c) Workspace. The Workspace is a destination to either process and prepare data for analytics applications (data engineering) or design and test own algorithms using machine learning and artificial intelligence tools (data science), a data factory so to speak [18]. Data Exchange or Marketplace is a bit like a data supermarket; it provides access to open data and some commercial offerings to complement or enrich own data [19]. Data scientists can pick data products off the shelf and make it available in Workspace. It also works the other way around: Data scientists can make their own data available as data products to others by putting them on the shelf of the data supermarket. As such, Data Exchange is an instant solution to overcoming data silos within companies, a company's own data supermarket: one store for all data for everybody. Furthermore, Data Exchange comes with centralized data governance, and policy and license enforcement. So, management will be delighted as it kills two birds with one stone: instant data sharing but with built-in compliance and risk management. Finally, the Connector, allows for data to remain locally and be exchanged peer-to-peer, while data transactions are tracked and logged for ease of billing and auditing.

3. Metadata

Metadata is a complex and complicated affair. Merriam-Webster defines metadata as “data that provides information about other data” [20]. Based on the conceptualization of a communication system in communication theory [21] this includes data about pretty much everything but the content of the actual message, for example, information about the sending and receiving nodes, the channel, frequency of messaging, etc. This data be very valuable. One example of such meta data in a communication scenario involving consumers is anonymized signaling data from the mobile networks of telecommunication providers. For example, Deutsche Telekom’s T-Systems offers such data commercially as Motion Data that can describe movement and traffic flows in public spaces, e.g., in pedestrian zones, on the road or in local traffic. It is easy to see how this data can be valuable for various applications, such as traffic planning for roads, retail location and even outdoor advertising, as Motion Data measures visitor numbers, traffic volume, and origination of traffic. This example is also useful to highlight one of the complications with metadata, the compliance with regulation. In case of Motion Data, the GDPR must be observed, and data must be anonymized and protected accordingly [22].

Another issue with metadata is the very standardization of it. Let’s consider another metadata example that many of us are probably very familiar with: Card catalogs of libraries, the information that can help a user find a particular publication. In order to find a book written and published elsewhere a common standard or language had to be agreed upon. This standard establishes a common understanding of the meaning or semantics of the data. With software automation or digital processing of data semantics, encoding is required to make data machine readable. One prominent initiative is the Semantic Web as an extension of the World Wide Web through standards set by the World Wide Web Consortium (W3C) [23].

3.a. Metadata in the MDS

In the MDS, metadata are defined in terms of Resource Description Framework (RDF), also the metadata broker supports this data representation [16]. Furthermore, it has been identified that data providers and consumers want information about the semantics-meaning of the data offered. If an existing semantic vocabulary is used, it should be part of the metadata. The semantic vocabularies used should be available in a standardized electronic format and a tool should be provided for the maintenance of these vocabularies. The different vocabularies and data schemes, used to represent the data offered, should be made available in a technical system that supports the MDS “vocabulary hub”. In the first stage it is planned to provide the vocabulary hub as a library of independent vocabularies. This library may contain external vocabularies defined by different organizations such as Datex II, Transmodel/NetEX, Traveller Information Services Association (TISA), C-ROADS, TN-ITS among others as well as vocabularies maintained within the MDS. It is planned to make RDF-based vocabularies available in this hub and to enable registered users to build new vocabularies in case there is no appropriate vocabulary available.

3.b. Connecting DIH to MDS

Data providers can access the DIH marketplace with secured identity and access management, and publish their own data offers by uploading only the metadata information to the central metadata storage.

These data offers can be accessed and traced via the DIH Metadata Broker (within DIH marketplace) while the raw data is transferred via federated manner (Connector). If there was only one data space, everything would work fine. But as numerous data platforms exist and data spaces emerge, data flows will be broken and users unsatisfied. Similar as organizations avoid single vendor lock-in, data space users do not want to be locked and limited by any single data space either.

Specifically, in our reference the data providers want to make their data offers on DIH marketplace also available in MDS, meanwhile data users can search in MDS Metadata Broker and find the relevant data offers which are published from DIH marketplace, and then directly develop their business concepts more efficiently and economically. Incorporating the heterogeneity of the data flow within or across data spaces, a trusted data exchange is designed and implemented to provide the business the full control on the data visibility, accessibility and sovereignty (cf. Figure 1).

For establishing such a trusted data exchange between DIH and MDS, a "Data Space Adapter" component is implemented with the following functionalities (cf. Figure 3): The Data Space Connector maps DIH metadata to IDS-conformant metadata structure (used in MDS), meanwhile provides a REST interface to fill the MDS Metadata Broker with the adapted metadata of data offers on DIH marketplace. The Publisher function will register MDS-accessible data assets on DIH as resources in MDS Connector as well as in MDS Metadata Broker. The Data Retriever function delegates the data assets requests from MDS connector to internal interfaces of DIH, and from there makes the desired data assets available.

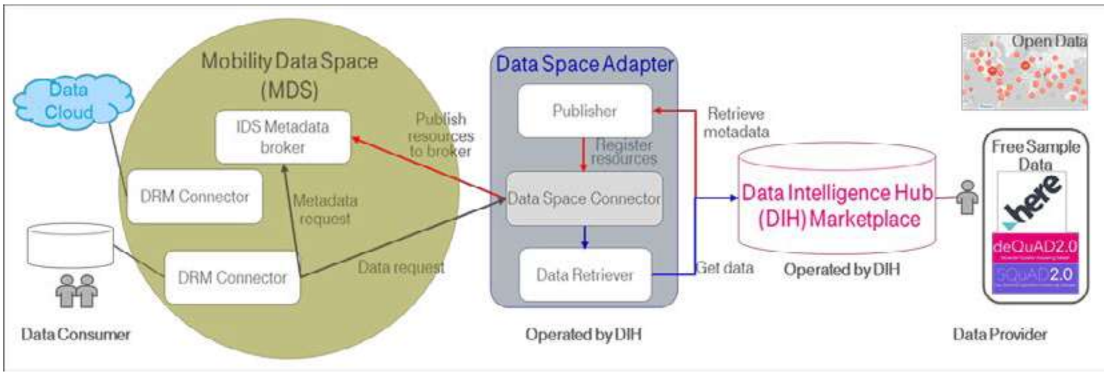

Figure 3 - Data Space Adapter as an on-ramp for data set onto MDS

3.c. Connecting NAP to MDS and to other NAPs

Collecting metadata from different platforms and making them searchable on other platforms is not a new topic. This is also called metadata harvesting and will be included in the Mobilithek from beginning on. Metadata from open data platforms (referring to data from the mobility sector) and from the MDS will be automatically integrated into the metadata broker.

With regard to the NAPs in Europe, the minimum implementation of a NAP, which has to be set up by each Member State of the European Union according to the ITS Directive, is a metadata repository. Representatives of four European NAPs have harmonized their metadata catalogues within the EU EIP project [24]. The first implementation of the Mobility Data Space has already used the currently available metadata catalogue and translated it into an RDF-compliant information model [16]. The harmonized metadata catalogue should then be transformed into an RDF-compliant DCAT Application Profile with the working title napDCAT-AP [25]. The harmonization efforts will now be built on a broader basis within the NAPCORE project including NAP operators from all EU Member States [26]. The usage of RDF-compliant metadata vocabularies within the NAPs will allow interoperability with respect to metadata between all NAPs and between NAPs and other data platforms in the mobility sector. Taking the example of Mobilithek, the German NAP, it aims to be integrated into MDS, i.e. to build a bridge between mainly public data at the NAP platform and the mainly private data within the MDS.

4. Trust

Trust in transactions is one of the most important pillars of a data sharing ecosystem. Trust enables the participation and fosters the sharing of data. There are many ways to achieve a trusted environment with technical means. Looking at the IDS reference architecture, trust is gained by the evaluation and certification of organizations as participants with focus on identity evaluation, as well as for the required software components. The evaluation is typically done by a dedicated company, playing the role of an *evaluation facility*, following a specific catalogue of criteria. The provision of the criteria is provided by the IDSA, which leads at least to an interoperability of trust between IDS-based data ecosystems.

In contrast, concepts like *verifiable credentials* focus on trust only for very specific information, rather than on a holistic evaluation of a whole organization, making the process leaner and more cost-effective. A combination of verifiable credentials and an IDS-based evaluation is also possible and could be complementary.

5. Additional interoperability aspects

Technical interoperability is a challenge as illustrated in Figure 4. Multiple, diverse systems must be connected to establish seamless data flows. However, early adoption of the IDS standard has so far mitigated this problem. In the example of MDS, Mobilithek and DIH (see Figure 2), all three efforts utilize IDS-compliant technology, including the Connector. The data space adapter (see Figure 3) is a good example to illustrate how the use of corresponding technology is beneficial. Using an IDS-compliant Connector, it can easily connect other data hubs into the MDS or allow third parties to publish data sets through the DIH on MDS.

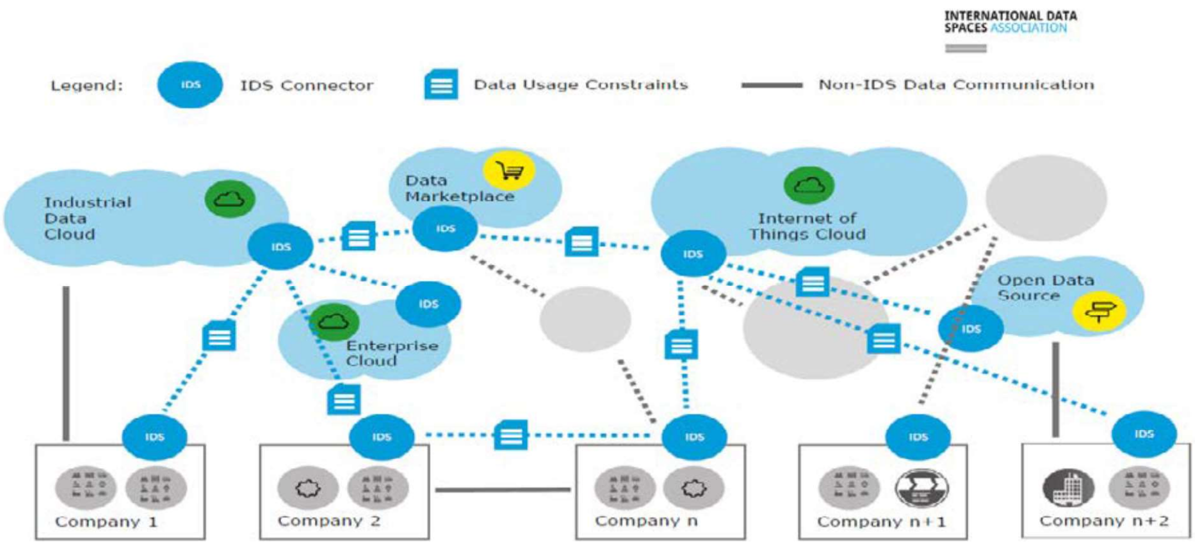

Figure 4 - Ecosystem of data spaces, clouds, platforms and marketplaces [14]

5.a. Organizational interoperability aspects

Besides technological there are organizational interoperability issues. Like a social or soccer club with club rules, a data space comes with rules. Users of a data space will be required to sign and agree to a legal contract specifying usage conditions before being admitted to the club. And rules may be independent of ownership type of a dataspace, which may be public, private, or public-private. While these governance and compliance rules as well as identity management greatly facilitate efficient interaction within a data space, they can present significant hurdles for interoperability between different data spaces and ecosystems.

5.b. Identity and access management

A trickier question is the management of technical identities to be ensured by an identity and access management (IAM) system across the platforms.

On the organizational level, additional topics need to be taken into consideration. These are general rules such as contractual agreements for using either platform (here used for dataspace, platform and marketplace) or the level of trust required to be participant of the platform e.g., specific certification requirements.

5.c. Platforms offering interoperable functionality

The platform itself, which acts as "platform-as-a-service" as described above in the DIH context, acts in the role as service provider according to the IDSA role concept [14]. In this concept, the service provider provides only the gateway for the data exchange. Before this gateway can be established, a comparison of the contracts and certification or trust level requirements for using each platform need to be compared. If the service provider becomes a participant of the MDS, he agrees to the rules which apply for using the platform. Assuming that the service provider, e.g. the Mobilithek, is not owner of the data, some considerations need to take place to establish the service provider only as a messenger between a data provider from one platform and a data consumer from another platform.

Technically, there are different options to establish the data exchange between the provider and the consumer. One is using one Connector for each participant. Another one would be to use one Connector as gateway between the platforms (see Figure 2). Independent from the technical solution, one Connector per participant or one Connector for the platform, the contract negotiation needs to take place between the data provider and the data consumer. The owner of the data, the data provider, needs to be named in the contract-data usage conditions – which are transferred between the two Connectors.

The service provider, who provides the technical gateway, has to offer additional services like the translation of the terms and conditions from text into machine-readable data usage policies to be transferred. Another prerequisite is that it should be clear in the metadata catalogue which participant is owning and in charge of a dataset presented through the platform.

5.d. Platforms offering intermediary data space services

To conform with the intense legislative agenda, data space service providers shall take part of the data space intermediation functions in the short term. This scenario is described in this session and is meant as an inspiration for platform providers, and further scenarios are also foreseeable in the future that a data space intermediary role might be developed standalone. Let's use the Data Space Adaptor build in DIH again as an example. DIH platform needs services to allow the platform participants to define the associate data license and policy directly during the data offer publishing process on DIH, and DIH ensures translating these data offer license and policy into usage policy. The platform, acting as a messenger, can communicate the data owner/provider and contractual party in the metadata with another data space, and therefore ensure the platform participants prepare the data offer once and publish in multiple data spaces. When a data provider/consumer uses DIH managed Connector as a Service (CaaS) without the platform, he will be a direct participant of the data space and can define and maintain the usage policies for their data directly in the Connector.

6. Governance Aspects for Data Spaces

While the described interoperability aspects are prerequisites for connecting data spaces and data platforms, it is essential that those prerequisites are kept alive on a longer term. Technologies, policy and also requirements of the participants will continuously evolve. For every interoperability aspect, ideas and proposals for governance mechanisms will be outlined in this chapter.

6.a. Metadata harmonization

Chapter 3 has shown that data needs to be described in a consistent manner with metadata in order to make it findable. Metadata vocabularies can give guidance on how to describe the data and define mandatory parts of metadata. Whereas this can slightly differ from one data space to another, a basis set of metadata should be always available. Depending on who is searching (machine or human being) different information might be interesting. Whereas for a manual search different categories and most searched keywords, just to name examples, might be helpful, for machine search this might not be relevant. As these requirements and guidelines for metadata can and will change over time, the challenge on governance level is not only to make the metadata of different data spaces and data platforms interoperable only once, but to keep the metadata vocabularies interoperable through the whole lifecycle. This can include continuous harmonization efforts and involvement of stakeholders

6.b. Trust

As described in chapter 4, IDSA provides a concept for allowing trustworthy data exchange of participating systems. The governance issue is to reach a common understanding and a common evaluation process to gain trust in transactions regardless of data space and ecosystem. The identification of participants and originations is the first crucial and important step and must be standardized among data space architectures and frameworks to achieve interoperability, at least on a European level. If further differentiation is needed, concepts like verifiable credentials could support domain specific or data space specific requirements.

7. Conclusions & Outlook

Data space concepts are designed to solve interoperability within. Now that first data spaces are arriving, interoperability across data spaces or mesh data spaces is becoming an issue. This paper focuses on this emerging topic. It uses two examples to start dissecting the issue. The examples are efforts to make the

Mobility Data Space initiated in Germany (MDS) interoperable with two existing data platforms, the German National Access Point for mobility data (NAP) and the Deutsche Telekom Data Intelligence Hub (DIH). Such data platforms can make emerging data spaces more attractive, enriching data offers. For one they can outright increase data offers (NAP, Mobilithek), for another they can make it easier for data providers to make data offers available (DIH with Data Space Adapter). Nevertheless, interoperability aspects are manifold with requirements on different levels. Making data from different platforms and data spaces discoverable is already possible on a technical level. Further interoperability aspects like the common trust anchor, organizational and legal issues are now being tackled and implemented in MDS, German NAP and DIH. This paper helps with transparency, because as data spaces are open ecosystems, the evolution of governance should be an open process as well. Whereas the newly founded non-profit organization for the MDS is currently setting up the operation environment of the MDS, the establishment of a sustainable governance framework will follow.

Acknowledgment

This work has been funded by the Federal Ministry for Digital and Transport (BMDV), a. o. by Modernity Fund (mFUND) under grant agreement 19F2083B ("Verknüpfung kommunaler, regionaler und nationaler Datenplattformen durch Data-Space-Konzepte sowie Veredelung und Verwertung als Mobilitätsdaten-Ökosystem - Mobility Data Space").

References

1. European Commission: A European Strategy for Data. COM(2020) 66.
2. Nagel, L., D. Lycklama (eds) (2021): Design Principles for Data Spaces – Position Paper, Dortmund. International Data Spaces Association. https://design-principles-for-data-spaces.org/
3. https://www.gartner.com/en/documents/4008708
4. https://www.forrester.com/allSearch?query=mesh&s=relevance&dateRange=365
5. https://www.oracle.com/a/ocom/docs/datamesh-ebook.pdf
6. European Union (2017). New European Interoperability Framework (EIF) – Promoting seamless services and data flows for European public administrations
7. https://www.go-fair.org/fair-principles/ (CC-BY 4.0)
8. https://mobility-dataspace.eu/
9. https://www.acatech.de/projekt/datenraum-mobilitaet/
10. https://www.mdm-portal.de/
11. https://www.here.com/platform
12. https://dih.telekom.net/

13. Acatech (2021). Vernetzter Verkehr: acatech initiiert Trägergesellschaft „DRM Datenraum Mobilität GmbH“ als Non-Profit-Organisation. https://www.acatech.de/allgemein/vernetzter-verkehr-acatech-gruendet-traegergesellschaft-drm-datenraum-mobilitaet-gmbh-als-non-profit-organisation/

14. International Data Spaces Association (2019), Reference Architecture Model Version 3.0, Dortmund. https://internationaldataspaces.org/use/reference-architecture/

15. Pretzsch, S., H. Drees, C. Lange, C. Mader, N. Petersen, L. Rittershaus (2017). Mobility Data Space – An Open and Decentral Ecosystem for Mobility Data. In Proceedings 25th ITS World Congress, Copenhagen. ERTICO (ITS Europe).

16. Drees, H., D. O. Kubitza, J. Lipp, S. Pretzsch, C. S. Langdon (2021). Mobility Data Space – First Implementation and Business Opportunities. In Proceedings 27th ITS World Congress, Hamburg, ERTICO (ITS Europe).

17. https://mobility-data.space

18. Schlueter Langdon, C., and R. Sikora (2020). Creating a Data Factory for Data Products. In: Lang, K.R., J.J. Xu et al. (eds). Smart Business: Technology and Data Enabled Innovative Business Models and Practices. Springer International Publishing. https://doi.org/10.1007/978-3-030-67781-7_5

19. Schlueter Langdon C. 2020. Dataspace Enabled Mobility. In: Mertens, C., and C. Schlueter Langdon (eds.). Data Move People Anthology (version 1.0), International Data Spaces Association, Berlin, Germany: 27-41

20. https://www.merriam-webster.com/dictionary/metadata

21. Shannon, C. E. (1948). A Mathematical Theory of Communication. The Bell Systems Technical Journal 27 (July, October): 379–423, 623–656

22. Art. 46 General Data Protection Regulation. https://gdpr-info.eu/

23. https://www.w3.org/standards/semanticweb/

24. Lubrich, P., et al. (2019). Coordinated Metadata Catalogue, EU EIP SA 4.6. https://www.its-platform.eu/wp-content/uploads/ITS-Platform/AchievementsDocuments/NAP/EU%20EIP_Coord.%20Metadata%20Catalogue_v2.0_191115.pdf

25. https://www.its-platform.eu/achievement/monitoring-harmonisation-of-naps/

26. http://napcore.eu/