The MobiSpaces Manifesto on Mobility Data Spaces

Christos Doulkeridis
University of Piraeus
Piraeus, Greece
cdoulk@unipi.gr

Ioannis Chrysakis
Netcompany-Intrasoft
Luxembourg
Ghent University
Belgium
KU Leuven
Belgium
Ioannis.Chrysakis@netcompany.com

Sophia Karagiorgou
UBITECH
Larnaca, Cyprus
skaragiorgou@ubitech.eu

Pavlos Kranas
LeanXcale
Madrid, Spain
pavlos@leanxcale.com

Georgios Makridis
University of Piraeus
Piraeus, Greece
gmakridis@unipi.gr

Yannis Theodoridis
University of Piraeus
Piraeus, Greece
ytheod@unipi.gr

Abstract

Data spaces consist of trusted frameworks that manage the entire data lifecycle, encompassing various data models, metadata descriptors, ontologies for semantic interpretation, and data services for accessing, processing, and analyzing data. Domain-specific data spaces are currently being designed and deployed in vertical sectors, following specifications and reference frameworks that enable interoperability and compatibility. One such vertical sector with high impact on economy is mobility. In this paper, we present our position towards mobility data spaces, highlighting the idiosyncrasy of mobility data, while presenting the MobiSpaces approach that encompasses services for mobility data management and advanced mobility analytics. By exploiting MobiSpaces, mobility data providers can break the barrier of participation in data spaces thereby sharing their dataset with less hurdle, whereas data consumers can find advanced data analysis tools readily applicable on mobility data sets to extract insights and discover mobility patterns.

CCS Concepts

• **Information systems → Data exchange**; **Spatial-temporal systems**.

Keywords

data spaces, data ecosystems, mobility data, mobility analytics

ACM Reference Format:
Christos Doulkeridis, Ioannis Chrysakis, Sophia Karagiorgou, Pavlos Kranas, Georgios Makridis, and Yannis Theodoridis. 2024. The MobiSpaces Manifesto on Mobility Data Spaces. In *4th Eclipse Security, AI, Architecture and Modelling Conference on Data Space (eSAAM 2024), October 22, 2024, Mainz, Germany*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3685651.3685654

eSAAM 2024, October 22, 2024, Mainz, Germany

ACM ISBN 979-8-4007-0984-5/24/10
https://doi.org/10.1145/3685651.3685654

1 Introduction

Data spaces [2, 17] comprise trusted frameworks for managing the complete lifecycle of data, covering various data models, metadata descriptors, ontologies for semantic meaning, as well as data services for accessing, processing and analysing data. Franklin et al. [7] originally introduced the concept of data spaces aiming to cover all data sources within an organization, irrespective of data model, data format or data location.

Recently, the concept of data spaces has been revived due to the observed situation worldwide that clearly indicates that companies and organizations mostly operate as "data silos" or "data islands", without well-established procedures to facilitate data discovery, trusted and standardized data sharing, exchange and interoperability, leading to a waste of resources due to unnecessary and repetitive data-related operations [3, 18].

As such, the concept of data spaces has similarities with data marketplaces, which are online transactional locations or stores that facilitate the buying and selling of data, according to Snowflake[1]. Besides the provision of a trusted infrastructure for data discovery, sharing and exchange, data spaces comprise fully decentralized infrastructures that promote interaction in a federated or peer-to-peer way. In turn, this enables the participants or actors of data spaces to take the roles of data provider and/or data consumer, possibly after negotiation and settlement for a specific monetary value. Apart from data exchange, data spaces support the provision and consumption of services, thereby offering a fully-fledged platform for building next-generation applications.

As the mobility domain is one of the fundamental pillars of the modern digital economy worldwide, the need for data spaces tailored to the mobility domain is imperative. The distinguishing features of mobility data spaces relate to the idiosyncrasy of manipulation of mobility data; distributed data acquisition, imprecise data recording of spatial data, error detection and cleaning of spatio-temporal records, semantic data representation, analysis and learning of mobility patterns are just a few of the rising challenges in

[1] https://www.snowflake.com/guides/what-data-marketplace

the mobility domain. Consequently, recent initiatives for mobility data spaces are trying to pave the way towards future mobility platforms, while also keeping up with modern research advances in mobility analytics [6] and mobility data science [15].

Towards this goal, MobiSpaces contributes to the design of future mobility data spaces by:

- offering data governance services for mobility data for data providers,
- supporting data consumers via advanced data analysis services,
- providing these services in a data spaces compliant way, leveraging data connectors and the concept of data-app,
- in a trusted environment that respects the architectural design of data spaces.

The remainder of this article is structured as follows. Section 2 provides a crisp overview of the concept of data spaces. Then, Section 3 briefly presents the vision of MobiSpaces for data spaces focused on mobility data and services, whereas Section 4 and 5 elaborate on the details for data providers and consumers respectively. Moreover, Section 6 focuses on the aspect of trust. Finally, Section 7 summarizes the main points and discusses future challenges in the domain.

2 Data Spaces: An Overview

2.1 Initiatives

The International Data Spaces Association (IDSA). It is a coalition comprised of more than 150 international organizations[2], which emerged in 2016 and has worked on the concept of data spaces and the principles that their design should follow in order to obtain value from data through sharing. The companies that comprise the IDSA represent dozens of industry sectors and are based in more than 20 countries around the world. As IDSA's mission is to drive the global digital economy, the main outcome of this work carried out by this coalition is to promote a Reference Architectural Model (RAM), called International Data Space (IDS) -RAM [3]. The aim of this model is to standardise data exchange in such a way that participants can obtain all possible value from their information without losing control over it. This IDS-RAM is characterized by an open architecture, which is reliable and federated for cross-sector data exchange, facilitating sovereignty and interoperability.

Gaia-X Gaia-X[4], first introduced by the German and French Ministries of Economics in October 2019, is an initiative that unites various stakeholders to facilitate data and service sharing. Gaia-X enables data and service sharing amongst participants through the use of Federation Services. It can also merge different ecosystems by agreeing on a common operational model, i.e., the Gaia-X Operational Model, which is based on the defined Gaia-X basic concepts or the Gaia-X Conceptual Model. This concept is facilitated by the use of different planes, specifically the usage plane for technical interoperability, the management plane for governance, and the trust plane, which is supported by the common Gaia-X Trust Framework[5]; This framework provides the necessary set of rules that define the minimum baseline to be part of the Gaia-X ecosystem.

The Data Spaces Business Alliance In September 2021, a collaboration was established between the Big Data Value Association, FIWARE Foundation, Gaia-X, and the IDS Association, resulting in the formation of the Data Spaces Business Alliance. The aim of this alliance was to create a common reference technology framework by converging existing architectures and models. The framework developed by this alliance enables businesses to securely share and exchange data, thus promoting innovation, growth, and competitiveness. The alliance unites a diverse range of businesses, IT providers, research institutions, and other interested stakeholders from different sectors, all dedicated to the development and utilization of data spaces.

Open DEI project OPEN DEI[6] is an EU-funded project, which aims to detect gaps, encourage synergies, support regional and national cooperation, and enhance communication among the Innovation Actions implementing the EU Digital Transformation strategy. The cross-Industry Digital Platforms federation of the OPEN DEI project provides useful insights to the most relevant work in the field of Reference Architecture for building data ecosystems (e.g., data spaces) to support the digital transformation journeys in the four sectors targeted by OPEN DEI (i.e., manufacturing, agriculture, energy, and healthcare). This initiative aims investigate a conceptual overview of the offered DEs to guide their development.

Data Spaces Support Center The Data Spaces Support Center (DSSC)[7] is an initiative by the European Commission under the Digital Europe Programme to support the development and implementation of data spaces across various sectors. It aims to facilitate the creation of common European data spaces, which is a part of the European strategy for data. This is accomplished mainly by providing essential components that support the creation of data spaces, utilizing a range of technologies. In addition to that, the DSSC offers guidance on matters such as conceptual modeling (e.g., taxonomy, glossary) and the landscape of standardization (e.g„ a compilation of standards).

iShare iSHARE[8] is an initiative that was launched with the goal of streamlining and standardizing data sharing, initially within the logistics sector. This initiative was set in motion by the Neutral Logistics Information Platform, a Netherlands-based foundation. The iSHARE project introduces a consistent set of agreements or norms for identification, authentication, and authorization. This allows entities within the logistics sector to share data in a more straightforward and secure manner, even with parties they have not previously interacted with. At the heart of this initiative is the creation of a trust framework that eases the process of data sharing among participants in the data ecosystem. The ultimate aim of iSHARE is to apply this framework along with existing initiatives (such as IDS, Gaia-X, FIWARE) to establish a solid European trust network for B2B data sharing.

[2]https://internationaldataspaces.org/we/members/
[3]https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4
[4]https://gaia-x.eu/
[5]https://docs.gaia-x.eu/policy-rules-committee/trust-framework/22.10/
[6]https://www.opendei.eu/.
[7]https://dssc.eu
[8]https://ishare.eu/

Mobility Data Space (MDS)[9] The Mobility Data Space (MDS) [19], launched in early 2022, represents an advance in intelligent traffic and mobility systems. It is an open, IDS-based data space that overcomes challenges in utilizing mobility data due to technical, legal, and economic constraints. MDS ensures the secure exchange and integration of sensitive mobility data across platforms. Key components of MDS include IDS Identity Provider, Metadata Broker, App Store, Clearing House, Vocabulary Provider, and Connectors, which collectively bolster its capability in managing and leveraging mobility data effectively.
PrepDSpace4Mobility[10] PrepDSpace4Mobility was a 12-month Coordination and Support Action (CSA) European Project aimed at advancing the common European mobility data space. The project involved mapping existing data ecosystems, identifying gaps and overlaps, and proposing common building blocks and governance frameworks found in current data space architectures. The project's outcomes include a comprehensive inventory of data ecosystems, including National Access Points for providing mobility data and relevant services, and the definition of the main pillars for a data space that will facilitate the deployment of the common European Mobility Data Space[11].
deployEMDS[12] deployEMDS is a follow-up project co-funded under the EU Digital Europe Programme, designed to establish the necessary framework for interlinking existing federated ecosystems. This will be achieved by integrating common technical infrastructure and governance mechanisms for the upcoming implementation of the common European Mobility Data Space. The initiative fosters a wide European ecosystem of data providers and users, promoting the adoption of shared building blocks. Sixteen real-life use cases from nine EU countries have contributed to developing innovative services and applications in this area. The ultimate goal of deployEMDS is to accelerate sustainable and smart mobility, thereby reducing transport emissions, by focusing on implementing its infrastructure across the nine project sites.

2.2 Basic Concepts – IDS

The participants in a data space are entities that are interested in sovereign data exchange in a trusted framework and in the use of services/apps that can be applied to shared data.

The IDS-RAM V.4.0 defines five (5) layers in the architecture: (i) the business layer, (ii) the functional layer, (iii) the information layer, (iv) the process layer and (v) the system layer. For the purposes of this document, the business layer is of main interest, which defines the main roles of participants and the patterns of their interactions.

Participants in data spaces may take one of a predefined set of *roles*: core participant, intermediary, software developer, or government body. Core participants are those that interact every time a data exchange takes place. These typically correspond either to the role of *data supplier* (i.e., data provider) or to the role of *data customer* (i.e., data consumer). On the other hand, intermediaries are platform roles that can be assumed only by trusted organizations and are important for the extended functionality of the data space

[9]https://mobility-dataspace.eu/
[10]https://mobilitydataspace-csa.eu/
[11]https://digital-strategy.ec.europa.eu/en/policies/mobility-data
[12]https://deployemds.eu/

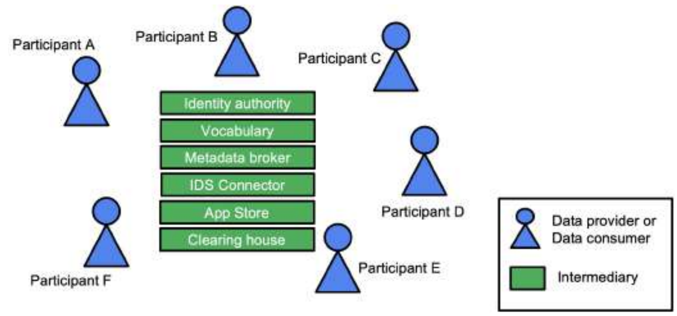


Figure 1: Abstract illustration of main roles in data spaces: data providers, data consumers and intermediaries.

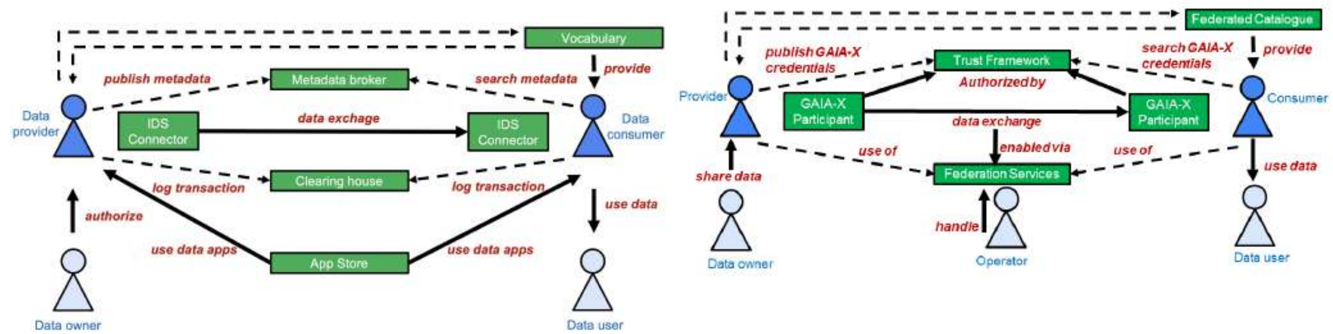
and they can be used by all data suppliers and customers. Based on IDS-RAM 4.0, the intermediaries are (depicted also in Figure 1):

- Identity authority: a service for identity management, ensuring security and preventing unauthorized access to data.
- Vocabulary: a service that manages and offers vocabularies, in the form of ontologies, data models and metadata. Vocabularies can serve as data annotations.
- Metadata broker: a service that maintains the metadata for the available data sets. It supports receiving metadata for data sets, but also providing metadata about data providers.
- IDS Connector: a software that is responsible for the actual data exchange between data supplier and data consumer (see [4] for a recent survey).
- App store: a service that enables to download algorithms (so-called data apps) to be deployed locally in the IDS Connector of the App consumer.
- Clearing house: an intermediary that provides clearing and settlement services for all financial and data exchange transactions.

It is fair to say that data spaces involve the above main roles, despite small deviations between leading initiatives such as IDS and Gaia-X in the definitions of concepts and prototype implementations.

Figure 2a depicts the interactions that take place in order to perform data exchange in IDS ecosystem.

1. **Resource and Offer Creation by Data Provider:** Data providers prepare data sets or data services, creating resources for sharing within the IDS framework.
2. **Publishing Offers and Data Discovery:** Resources are transformed into offers with metadata and usage terms, cataloged by the IDS Metadata Broker for discovery.
3. **Identification and Authentication:** Managed by the IDS Identity Provider, this step verifies and authorizes entities, securing the network's integrity.
4. **Negotiation and Agreement:** IDS Connectors facilitate the negotiation of data usage terms, adhering to predefined policies and contracts. During this phase, the IDS Clearing House is instrumental. It aids in negotiating terms of data usage, including access rights and duration, ensuring these negotiations adhere to established policies or contracts.



(a) IDS ecosystem

(b) Gaia-X ecosystem

Figure 2: Interactions between data providers/consumers and intermediaries in data spaces.

(5) **Data Loading and Transfer:** Data is transferred securely, with IDS Connectors ensuring proper format, encryption, and integrity.
(6) **Post-Transfer Processing and Compliance:** Post-transfer, data is processed per the recipient's needs and compliance with data usage control policies is enforced.

2.3 Basic Concepts – Gaia-X

The Gaia-X suggests mainly two ecosystems: the Data ecosystem and the Infrastructure ecosystem. Both share the elements coming from the Gaia-X architecture: Gaia-X Conceptual Model, Gaia-X Operating Model, Federation Services, and Gaia-X Trust Framework [8, 9]. The Gaia-X Conceptual Model provides a common understanding on the basic concepts used within Gaia-X[13] The fundamental element in this model is the "Gaia-X participant" representing an entity within a domain with distinct roles:

- Provider: delivers Resources and Services in the Gaia-X ecosystem, often referring to a Data provider.
- Consumer: searches for Service Offerings and consumes Service Instances defined within the Gaia-X data space, akin to a Data Consumer.
- Operator (previously known as Federator [8]): Oversees at least one Federation Service, functioning as an Intermediary.

As we can see, these roles are similar to the ones that identified by IDS. The *Gaia-X Operating Model* is introduced to manage the involvement of various participants, According to this model, a "*Trust Anchor*" is a designation for entities considered trustworthy within the data ecosystem. To become a Trust Anchor, participants follow a standardized nomination process facilitated by labeling mechanisms established by the Gaia-X association. Within the Gaia-X ecosystem, *Federation Services* play a vital role in enhancing data sharing scenarios. They achieve this by providing federating catalogues, allowing Consumers to discover optimal offerings and monitor them for relevant changes. For Producers, these services enable the promotion of offerings while maintaining full control over visibility and privacy levels.

[13]https://docs.gaia-x.eu/technical-committee/architecture-document/23.10/gx_conceptual_model

The *Gaia-X Trust Framework* encompasses all necessary components to ensure Gaia-X compliance. Key elements in this framework include *Gaia-X Credentials* (previously referred to as Self-Descriptions [8]), which are machine-readable files describing all entities. These credentials undergo validation against a defined schema, utilizing descriptions following the W3C Verifiable Credentials Data Model Standard[14]. They can also function as contracts, binding legal agreements between Gaia-X services.

Figure 2b illustrates the interaction process required for data exchange within the Gaia-X ecosystem. In this system, participants of Gaia-X are enabled to share data once they receive authorization from the Trust Framework. Both providers and consumers exchange Gaia-X credentials, also known as self-descriptions, which are utilized for authorization and data discovery. The process of data discovery is further amplified by Federated Catalogues, which provide information about the available services within Gaia-X Federation Services. Essentially, these Federation Services serve as the technical foundation for all the functionalities offered by the Gaia-X ecosystem and these are handled by specified Operators as participants of the data space. Examples of Federation Services include data exchange services and identity and access management services.

2.4 Minimum Viable Data Space (MVDS)

The MVDS is a tiny deployment of a data space that contains only a subset of critical components for secure and sovereign data exchange [15]. In practical terms, a typical implementation of a MVDS consists of: (i) two or more Connectors, (ii) the Certificate Authority granting X.509 certificates, and (iii) the Dynamic Attributes Provisioning Service (DAPS) to handle dynamic attributes and manage dynamic access tokens.

The MVDS approach aims to simplify the process of initiating a data space by focusing on essential features that are necessary for it to be operational for secure data exchange (see also [1]). This approach is particularly beneficial for experimenters and developers, as it shortens the implementation time by omitting complex details that could delay the initial release.

[14]https://www.w3.org/TR/vc-data-model/
[15]https://docs.internationaldataspaces.org/knowledge-base/mvds

In the context of MVDS, *Connectors* are crucial components that enable seamless interaction and data exchange between different systems and sources. They integrate data from various formats, locations, and models, forming the backbone of the MVDS architecture. By bridging diverse data environments, connectors support the key objectives of MVDS, such as data discovery, sharing, and interoperability. Their robust functionality allows organizations to maximize their data assets, driving innovation, efficiency, and collaboration. The IDSA regularly publishes the Data Connector Report[16], which provides updates on IDS connectors' development and usage. Notable examples include the Eclipse Connector[17] and the Dataspace Connector[18], which are widely implemented across various projects. We utilize the Dataspace Connector within MobiSpaces to establish a dedicated mobility data space.

3 The MobiSpaces Approach for Mobility Data Spaces

MobiSpaces[19] is a Horizon Europe research and innovation project that aims to deliver an innovative, effective, robust, and green ecosystem for the entire lifecycle of mobility data [5, 10]. It relies on the results of IDS and Gaia-X and is compliant with their high-level design and architecture. However, MobiSpaces develops innovative solutions that go a step further from existing data spaces initiatives, by incorporating methods, algorithms and techniques for mobility data management, mobility pattern discovery and machine learning. To demonstrate the versatility of MobiSpaces two mobility domains are considered: the urban domain and the maritime domain.

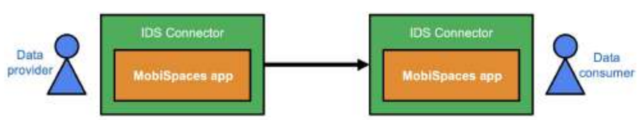


Figure 3: The MobiSpaces app and data connectors.

3.1 The Data-App Connectivity with MobiSpaces Ecosystem

In the realm of the MobiSpaces project, a novel data application (app) plays a crucial role in bridging the gap between the extensive mobility data repository of MobiSpaces and its diverse user base. This app, leveraging the state-of-the-art data spaces Connector technology (i.e., IDS Connectors) inherent to the MobiSpaces framework, facilitates the efficient, secure retrieval and provision of mobility data, as depicted in Figure 3. Designed with a user-centric approach, the app embodies an interface that simplifies the intricacies of data querying and analytical processes, tailored to suit the multifaceted needs of its users.

[16]https://internationaldataspaces.org/data-connector-report/
[17]https://github.com/eclipse-edc/Connector
[18]https://github.com/FraunhoferISST/DataspaceConnector
[19]https://mobispaces.eu/

- For data providers, the app offers an array of functionalities, encompassing data upload, management, and quality assurance, ensuring adherence to the FAIR principles (cf.Section 4).
- Conversely, for data consumers, it is equipped with advanced analytical features, enabling the derivation of meaningful insights through sophisticated data processing methodologies, inclusive of AI-driven analytics and comprehensive visualization tools (cf.Section 5).

This symbiotic integration of the app within the MobiSpaces ecosystem not only augments the accessibility and utility of mobility data but also fosters an environment conducive to seamless and secure data sharing. This is instrumental in driving innovation across urban and maritime mobility sectors. Thus, the app epitomizes the essence of the MobiSpaces project's vision: to revolutionise the landscape of mobility data spaces through a harmonious blend of technical prowess and user-focused design.

In the following, we elaborate on the two ways MobiSpaces can assist in the establishment of a data space for mobility data: (a) by empowering the role of data provider, and (b) by empowering the role of data consumer.

4 Empowering the Role of Data Provider

Consider a data owner that operates on curation of mobility data. This entails data acquisition, data manipulation, error detection, cleaning and eventually storing this curated data in a desired format.

To provide access to its data in a secure and trustworthy way to external parties, the data owner decides to participate in a data space. ***The* MobiSpaces *technology can assist the data owner to bridge the gap between its internal data management architecture with the architecture advocated in data spaces*.** This is achieved by means of MobiSpaces's *Data Governance Framework* and *AI-based Data Operations Toolbox*.

4.1 The Data Governance Framework

Data governance operations are essential for modern data spaces, in order to promote trusted data exchange and ensure reliability. This is also the case for mobility data which are prevalent in everyday operations, from route discovery to fleet monitoring.

The Data Governance Framework in MobiSpaces is an assembly of *data services* that operate both in standalone mode as well as by interacting with each other. These data services include error detection and cleaning, semantic representation and modeling, making data respect the FAIR principles [12, 28], data interlinking and data provenance. Basically, the Data Governance Framework supports data owners of mobility data with functionality related to data curation and data preparation, thus facilitating the participation of a data owner in a data space by taking the role of a data provider.

4.2 The AI-based Data Operations Toolbox

MobiSpaces offers efficient and scalable tools for querying mobility data, supporting a wide range of diverse scenarios. The AI-based Data Operations Toolbox consists of a *batch layer* and a *real-time layer*. Regarding the batch layer, MobiSpaces adopts a declarative querying approach ("SQL on everything") over all data sources, also popularized lately by Presto [24]. Users can express complex data processing tasks either by means of workflows (via a workflow builder) or by using SQL. Interestingly, SQL-based queries are supported for different types of storage of mobility data: raw or minimally-processed files, relational data, non-relational data, mobility data, as well as for encrypted data. As regards the real-time layer, this involves online data aggregation (e.g., for data compression) and a library of efficient mobility operators, which can be deployed at the Edge and support both data ingestion and in-situ querying.

Moreover, the Toolbox contains *infrastructure services*. A core component of these is the *Intelligent Resource Allocator*, which optimizes the allocation of resources based on various constraints provided as input. Resource allocation is modeled as a constrained optimization problem using this input, and assigning resources for the orchestration by the *Execution Orchestrator*. The target of deployment is based on Kubernetes clusters. For interaction with the end-user, a *Workflow Builder* provides a graphical interface that allows users to create complex, AI-based batch and ETL[20] workflows through a declarative methodology and a highly customizable, easy-to-interact canvas-and-palette interface. The workflow description is exported in .yml format. Then, the Intelligent Resource Allocator produces the optimized target description in the form of a decorated .yml-manifest. Thus, the interaction of these components allows the workload to be meticulously planned and distributed to the right execution environment, considering (i) the availability of both cloud- and edge-side resources, (ii) data access policies, and (iii) their location and availability based on mobility.

4.3 MobiSpaces Services for Data Providers

4.3.1 FAIRification. Before making a data set available in the data space, it is important to ensure that data is compliant with FAIR principles [28]. For this purpose, MobiSpaces provides support for semantic representation via a domain-specific ontology for mobility data. The ontology is an evolving result from previous EU research projects and has been applied to maritime and aviation mobility data [21, 27]. The ontology concepts form the *Vocabulary* for the data space.

In addition, MobiSpaces offers a tool (RDF-Gen) for semantic data transformation [22, 23] in accordance with the conceptualization of the domain offered by the ontology. Essentially, RDF-Gen accesses data from various sources (text/csv files, XML, JSON, binary files, database, etc.) and transforms it in RDF. In technical terms, the transformation is achieved by using *triple templates*, i.e., templates of RDF triples that are compliant with the underlying ontology. As the input data is accessed record-by-record, the values of each record are extracted, processed, and they populate the variables in the template. For example, consider a record describing a vessel named *HM196 FRU FJORD*, with identifier (MMSI code) *219001682*, registered in *Denmark*, with breadth *5m* and length *20m*, of type *fishing*. Then, the following triples template using TTL notation[21] can be used to generate RDF triples (following the schema of our ontology):

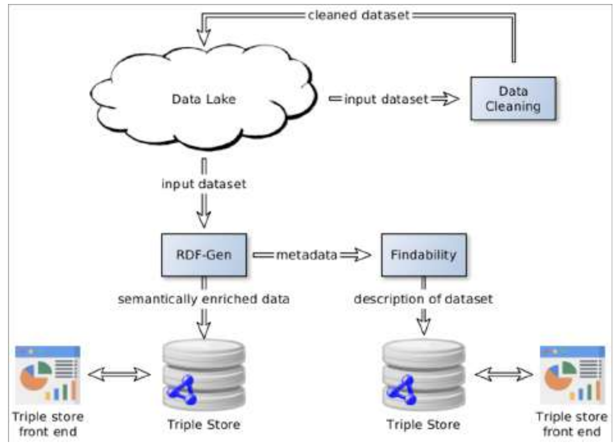


Figure 4: A workflow for FAIRification of mobility data.

```
getVessel(?mmsi) a getVesselType(?type);
:hasCode asString(?mmsi); :hasLength asString(?len);
:hasBreadth asString(?width); :hasName asString(?name);
 :registeredAt getCountry(?country).
```

The variables in the template (starting with the character ?) are bound with the record's values and they can be used directly or after having undergone some custom processing. For example, function *getVessel()* takes the vessel's identifier *219001682* and produces a URI that contains this value. We have a rich and extensible palette of processing functions, which allow users to simply edit their own triples template based on their data and ontology, without the need of recompiling RDF-Gen. In this way, every data object is associated with a global unique identifier (URI), making the data *accessible* and *interoperable*. In turn, the RDF data can be interlinked with other open linked data sets. In this way, we can obtain enriched semantic representations of data.

Another interesting contribution of MobiSpaces concerns the *findability* of data sets. During the process of parsing and transforming a data set to RDF, we automatically construct and maintain various metadata, which are typically used when searching mobility data sets. Examples include the geographical box that encloses all spatial positions, the timespan of the records in the data set, types of moving objects found in the data set, etc. This metadata is collected for each data set automatically and can be used by the *Broker* of the data space.

Figure 4 shows the combination of our data services into a FAIRification workflow for mobility data that can be used at the data provider side. The input data is initially stored on-premises, for example in a data lake at the data provider's side. To make this data publishable in a data space, MobiSpaces offers data cleaning (see Section 4.3.2), followed by data transformation in RDF (via RDF-Gen) and making the RDF data available in a RDF triple store. Moreover, during the data transformation, we compute useful metadata descriptors that are passed over to the findability component to serve as searchable parameters that are generated without human intervention.

[20] Extract-transform-load (ETL) is the process of extracting data from one or multiple sources, transforming into an appropriate format (possibly also applying cleaning, normalization, etc.) and eventually loading the data in a database.
[21] https://en.wikipedia.org/wiki/Turtle_(syntax)

MobiSpaces offers FAIRification data services that transform raw data into *findable*, *accessible* and *interoperable*, by using an RDF-based representations that can be used for data interlinking with other sources and eventually for extracting data in various other shareable formats. Moreover, for each data set, metadata are automatically generated that support content-based data discovery.

4.3.2 Mobility Data Cleaning and Error Detection. GPS data typically contains noise due to limitations of mobile sensor technology and due to other external factors. For example, the accuracy of GPS devices is known to be a couple of meters, while various obstacles such as buildings or bridges may have a negative effect on this. In the sea, similar issues exist as large vessels are required to use the Automatic Identification System (AIS) to report their position, and AIS data is noisy. Oftentimes, errors are introduced during the data acquisition process (e.g., longitude and latitude columns are reversed), data transformation (e.g., date-time conversion), or data filtering (e.g., spatial area clipping may remove GPS positions and lead to infeasible trajectories).

MobiSpaces includes an error detection and cleaning mechanism that is tailored for mobility data, aiming at error-free data sets that have higher value for training upstream machine learning models. This data service is useful for mobility data providers, as it contains domain-specific cleaning methods for spatial and mobility data that typically show superior performance compared to generic cleaning and error detection techniques for tabular data.

MobiSpaces offers error detection and cleaning data services that results in more accurate and reliable data, thus increasing the value of shared data sets.

4.3.3 Data Provenance. Tracking the provenance of data sets is very important for capturing all intermediate steps in data processing pipelines and machine learning workflows that mainly operate on data operations. In this way, all changes on data sets that were caused by applications can be recorded, allowing to query the lineage of data. In practice, the provenance mechanism supports querying the history of a data set that has been successfully used for training an accurate prediction model, so as to identify which processing steps (e.g., data transformation, cleaning, data normalization, anomaly/error detection and correction, missing value replacement, etc.) induced the highest increase in performance.

In MobiSpaces, we have built a REST API service that accepts POST requests from applications that operate on an input data set and produce an output data set. All data set instances are treated as resources and are associated with unique resource identifiers. The information about modifications and changes is captured and represented using the PROV-O ontology[22], a conceptualization targeted for tracking provenance of resources. In this way, the provenance information is represented in RDF and can be stored in an RDF store, allowing declarative querying using SPARQL.

MobiSpaces provides a data provenance mechanism that can track modifications on mobility data sets and also offers a declarative query interface for querying the provenance of any data set registered in the data provider side.

[22] https://www.w3.org/TR/prov-o/

4.3.4 Mobility-aware Edge Processing. Mobility data generation is performed in a decentralized and distributed way, by means of various sensors that track the mobility of objects. Typically, this movement data is transferred to the cloud, where data collection and data preparation methods are applied, before it is stored for further processing and analysis.

We advocate that mobility data spaces could exploit edge computing technology to optimize data-related operations. Several data curation processes (cleaning, anonymization, aggregation, compression, etc.) can be moved to the edge, in order to be applied in an online manner near to the sources of data generation. In turn, this offers many advantages: energy savings due to reduced network traffic and decentralized processing, increased data privacy due to data anonymization near the sources, optimized allocation of resources according to mobility patterns.

To this end, MobiSpaces offers tools to data owners for edge data management leading to optimized deployment of data acquisition services and optimized resource allocation. This is a largely neglected issue in data spaces, as they focus on subsequent steps and mainly in secure data sharing and standardized data exchange. However, in the mobility domain, an equally important problem concerns decentralized data acquisition, and MobiSpaces tries to address this emerging requirement.

MobiSpaces includes edge technologies that can be readily deployed on edge devices near to the sources of data generation, including an orchestrator for multi-cloud settings and an intelligent resource allocator for optimized allocation of edge and cloud resources.

5 Empowering the Role of Data Consumer

At the data consumer side, MobiSpaces ***offers a set of application services for the analysis of mobility data***. This is mainly provided by means of the *Edge Analytics Suite*, which offers diverse functionalities: distributed edge analytics, federated learning for mobility data, explainable artificial intelligence methods for the interpretation of mobility models, and visual analytics empowered by interactive map-based visualizations.

5.1 The Edge Analytics Suite

The Edge Analytics Suite consists of two main decentralized components that provide dynamic edge analytics algorithms and methods for federated learning over trajectory data. In addition, an eXplainable AI (XAI) component is included, which offers interpretability and explainability for the models learnt by the decentralized components. Last, but not least, a visual analytics component closes the loop and enables human interaction with the results of data analytics, the learned models and their interpretation, by means of interactive visualizations that can assist a domain expert in discovering meaningful patterns of movement.

5.2 MobiSpaces Services for Data Consumers

At the side of data consumer, the main need is to apply data analytics and machine learning methods to mobility data sets, in order to discover hidden mobility patterns that may lead to invaluable business insights. To this end, MobiSpaces offers various mobility-specific data analytics methods and techniques that provide data consumers with a rich repertoire of models.

From a technical standpoint, data consumers in the MobiSpaces platform can leverage advanced data analytics and machine learning methods to uncover latent mobility patterns, leading to significant business insights. MobiSpaces offers a diverse array of mobility-specific data analytics methods and techniques, equipping data consumers with a rich repertoire of analytical models.

One of MobiSpaces's primary offerings is a suite of machine learning (ML) algorithms specifically designed for mobility data. These algorithms cover essential tasks such as trajectory cleaning and segmentation, trajectory forecasting, and anomaly detection. These application services are seamlessly integrated within the data space, allowing data consumers to access and utilize these tools efficiently. Furthermore, MobiSpaces supports both centralized algorithms and those tailored for distributed settings, such as edge analytics and federated learning, ensuring flexibility and scalability across various operational environments. MobiSpaces integrates federated AI to address the complexity of trajectory data, enabling the creation of distributed movement models in a cross-device federated learning setting. This involves multiple mobile devices computing distributed movement models that contribute to updating a global model. Federated learning in MobiSpaces encompasses classification and regression tasks as well as advanced deep learning [11, 26].

Additionally, MobiSpaces emphasizes the explainability of trained ML models through advanced XAI techniques tailored for the mobility domain. These techniques are integrated into an XAI framework [13]. Applying XAI models to the mobility data and ML models poses a challenge because most XAI methods expect 2D arrays (samples, features), while mobility data are represented as 3D inputs (samples, timesteps, features). To address this, we created a wrapper function around the AI model's prediction function to reshape the 2D input from XAI method to the 3D input format the AI model. The features consistently appearing as top contributions indicate which aspects of the data the model focuses on. In a flattened time series, these are specific timesteps (Fig. 5). To enhance the accessibility and practical application of XAI models in Vessel Route Forecasting (VRF), MobiSpaces offers a GPT-based model using OpenAI technology based on the solution of [14]. This model is designed to bridge the gap between complex AI outputs and the end users' understanding, thereby promoting transparency and trust. By leveraging the NLP capabilities of GPT, the model translates technical insights and model interpretations into clear, actionable recommendations. This approach ensures that even those with limited technical expertise can effectively utilize the advanced capabilities of XAI models. An example of this output can be seen in Fig. 6. These methods enable data consumers to build interpretable ML models and understand the decisions made by these models.

Moreover, MobiSpaces includes sophisticated visual analytics applications that enhance human-computer interaction through interactive visualizations. These tools enable users to engage deeply with the data, exploring and interpreting mobility patterns intuitively and comprehensively. The Privacy-Aware Visual Analytics (PAVA) module, for example, supports data retrieval, transformation, and visualization while ensuring data privacy and security. PAVA

Figure 5: LIME-based highlighted trajectory. The historical trajectory is depicted and the most important parts of the historical route are highlighted based on the LIME weights.

allows users to generate personalized visualizations, such as heat maps and trajectory overlays, which are crucial for understanding the insights derived from mobility data.

In line with the integration of advanced MLOps (Machine Learning Operations) workflows MobiSpaces leverages a data connector/consumer approach to streamline the integration and deployment. By integrating MLOps with IDS connectors, MobiSpaces revolutionizes the handling of data in the ML pipeline, ensuring that the entire process—from data collection and processing to model training,—is efficient and secure.

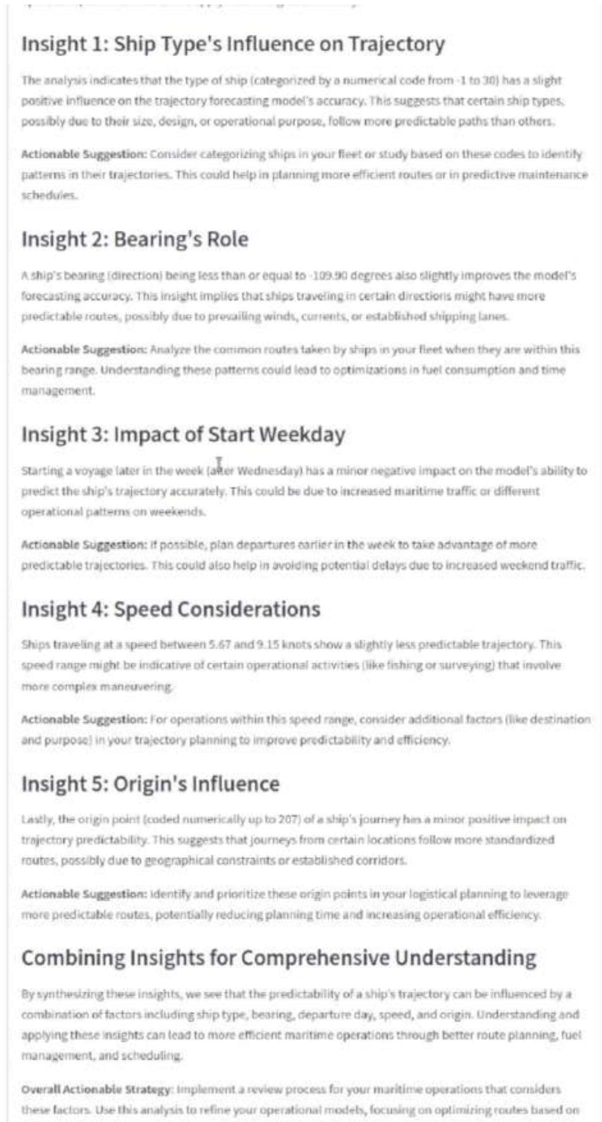
MobiSpaces offers advanced algorithms for data analytics and machine learning for mobility data sets, that can be readily applied on maritime and urban data. Moreover, it provides eXplainable AI methods for models trained on mobility data.

6 Establishing Trust in MobiSpaces

The Internet architecture lacks the essential concept of trust, which is how to build an ecosystem of trusted entities required to consume services and interact with each other in a trustworthy manner. Trust is the basis for data spaces, where each participant is evaluated and certified before being granted access to the trusted business ecosystem. This is also necessary for mobility data spaces to support federated and non-monopolised Internet services with many concurrent cloud providers and edge nodes both acting as consumers and producers of data.

At the same time, the European Union[23] is heavily investing in the Decentralised Identity Framework ESSIF (European Self Sovereign Identity Framework) [20] used for next-generation digitalisation (Web3[24]) in the public and private sector. Under the same perspective, DG DIGIT incubates the concept of Trust and Identity in the support of EU Data Spaces. As part of data governance structures, EU Data Spaces[25] should also determine the rights concerning access to and processing of the data.

[23] https://www.eesc.europa.eu/sites/default/files/files/1._panel_-_daniel_du_seuil.pdf
[24] https://web3.foundation/
[25] https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces



Insight 1: Ship Type's Influence on Trajectory

The analysis indicates that the type of ship (categorized by a numerical code from -1 to 30) has a slight positive influence on the trajectory forecasting model's accuracy. This suggests that certain ship types, possibly due to their size, design, or operational purpose, follow more predictable paths than others.

Actionable Suggestion: Consider categorizing ships in your fleet or study based on these codes to identify patterns in their trajectories. This could help in planning more efficient routes or in predictive maintenance schedules.

Insight 2: Bearing's Role

A ship's bearing (direction) being less than or equal to -109.90 degrees also slightly improves the model's forecasting accuracy. This insight implies that ships traveling in certain directions might have more predictable routes, possibly due to prevailing winds, currents, or established shipping lanes.

Actionable Suggestion: Analyze the common routes taken by ships in your fleet when they are within this bearing range. Understanding these patterns could lead to optimizations in fuel consumption and time management.

Insight 3: Impact of Start Weekday

Starting a voyage later in the week (after Wednesday) has a minor negative impact on the model's ability to predict the ship's trajectory accurately. This could be due to increased maritime traffic or different operational patterns on weekends.

Actionable Suggestion: If possible, plan departures earlier in the week to take advantage of more predictable trajectories. This could also help in avoiding potential delays due to increased weekend traffic.

Insight 4: Speed Considerations

Ships traveling at a speed between 5.67 and 9.15 knots show a slightly less predictable trajectory. This speed range might be indicative of certain operational activities (like fishing or surveying) that involve more complex maneuvering.

Actionable Suggestion: For operations within this speed range, consider additional factors (like destination and purpose) in your trajectory planning to improve predictability and efficiency.

Insight 5: Origin's Influence

Lastly, the origin point (coded numerically up to 207) of a ship's journey has a minor positive impact on trajectory predictability. This suggests that journeys from certain locations follow more standardized routes, possibly due to geographical constraints or established corridors.

Actionable Suggestion: Identify and prioritize these origin points in your logistical planning to leverage more predictable routes, potentially reducing planning time and increasing operational efficiency.

Combining Insights for Comprehensive Understanding

By synthesizing these insights, we see that the predictability of a ship's trajectory can be influenced by a combination of factors including ship type, bearing, departure day, speed, and origin. Understanding and applying these insights can lead to more efficient maritime operations through better route planning, fuel management, and scheduling.

Overall Actionable Strategy: Implement a review process for your maritime operations that considers these factors. Use this analysis to refine your operational models, focusing on optimizing routes based on

Figure 6: Example of GPT output for human-centric XAI.

The "Trust" concept in IDS covers three aspects, i.e., roles, identity management, and user certification, which are complemented by governance aspects [16]. Also, the "Security and data sovereignty" concept contains four major aspects, i.e., authentication and authorization; usage policies and usage enforcement; trustworthy communication and security by design; and technical certification.

In Gaia-X, "Trust" is enabled by applying the principles of Self-Sovereign Identity (SSI)[26]. Self-Sovereign Identity ensures secure and trustworthy digitalisation in the decentralized ecosystem that Gaia-X aims to be [25]. Without the need for a conventional central Identity Management System (IdM), the Self-Sovereign Identity concept enables any subject, such as a person, organization or even a machine, to manage their digital identities and associated credentials like membership cards, certificates or Self-Descriptions in a self-sovereign manner. Crypto standards and Self-Sovereign Identity standards, in combination with Web3-compliant technology components, enable the Gaia-X ecosystem to gain the required level of trust without the need for centrally hosted and controlled Identity Providers (IDPs). In the Gaia-X ecosystem, the SSI concept is activated through the use of open W3C (World Wide Web Consortium) specifications such as the Decentralised Identifier (DID) and the Verifiable Credential (VC) data model as well as open-source projects including Hyperledger Aries, ESSIF (European Self-Sovereign Identity Framework) or IDUnion[27].

Figure 7: ABAC-based Trust in MobiSpaces.

In MobiSpaces, we elaborate on the trust and security concepts using different Attribute-based Access Control (ABAC) mechanisms (Figure 7). These mechanisms enable to enforce context-aware (e.g., based on location, time, etc.) and multiple levels of trust and security per user role, service and Application Programming Interface (API). This way, also organizations with limited resources and technical means are able to participate (at least as Data Consumers). To the extent possible, existing standards have been leveraged in the development of the ABAC mechanisms of MobiSpaces. From a design perspective, we address four key trust and security concepts, which are: *identity management* enriched with attributes for location aware edge services, *trust management*, *data access control*, and *data usage control*. The aim of the ABAC mechanisms of MobiSpaces is to combine existing, reliable approaches in a useful and meaningful way, and bridge gaps where necessary. The developed mechanisms support the evaluation of attributes of the subject, attributes of the object, and the formal relationship or access control rule or policy defining the allowable operations for subject-object attribute combinations. The mechanisms are supported by a Single Sign-On (SSO) solution deployed through the SSO as a Service (SSOaaS) paradigm. We are exposing a highly configurable API through a DevOps handler that enables users, applications and services drive the underlying layer like IaaS for Infrastructure or PaaS for Platform through the necessary configuration in a transparent but agnostic manner to the triggering process. SSOaaS in our context means to provide the ability for an application to manage authorizations and choose user attributes to set. The DevOps concept supports the combination of practices and tools designed to increase MobiSpaces's ability to deliver applications and services faster than traditional software development processes.

The ABAC mechanisms in MobiSpaces rely upon the assignment of subject attributes to subjects and object attributes to objects,

[26] https://gaia-x.eu/wp-content/uploads/2022/06/SSI_White_Paper_Design_Final_EN.pdf

[27] https://idunion.org/projekt/?lang=en

and the development of policies that describe the access rules for each object to the underlying environment. Each object within the environment can be tagged or assigned specific object attributes that describe the object. The advantage of this mechanism is its simplicity; the service authenticates the end user on one designated platform, application, service, or portal, enabling her to then use a variety of services without having to log in and out each time via an SSOaaS API. The SSOaaS paradigm avoids multiplication of passwords and increase the overall IT security. The ABAC mechanisms in MobiSpaces have been configured to support different Client applications, Users and policies at the backend. At the same time, the SSOaaS API has been exposed only through credentials, access tokens and secrets to secure access only to designated end users services under specific conditions dictated by the supported policies. The user is established as a subject within the environment by an administrator, and the user's characteristics are captured as subject attributes. This subject has a name, a role, and an organisation affiliation. Other subject attributes may include user status, and her minimum metadata (e.g., first name, last name, email, etc.). When new users arrive, existing users leave or characteristics of the subjects undergo modifications, it is necessary to update these subject attributes accordingly. Every object within the environment must have at least one policy that defines the access rules (i.e., read / view, write / update / modify, delete) for the object. This policy is normally derived from the rights we need to assign to each specific object for a set of users with different operations over the objects.

MobiSpaces establishes trust in federated and centralised mobility data spaces through ABAC mechanisms which have been developed to grant access based on users' location, role, the time of day, the edge device being used, the resource in question, and the desired action. These are all the attributes necessary to enforce secure authorization dynamically and in real-time.

7 Summary and Outlook

MobiSpaces is a research project that aims at producing innovative research technology for future mobility data spaces. It offers a wide variety of data services that empower the roles of data provider and data consumer in data spaces. Consequently, by adopting the MobiSpaces perspective, data owners can embark in mobility data spaces more easily and can exploit multiple ready-to-go tools and techniques for mobility data to harness the merits of data spaces.

Despite the recent interest in data spaces, their wide adoption and success in the future is subject to several factors. Domain-specific data spaces, such as mobility data spaces, need to offer assistive technologies that are compliant with data spaces architectures but also act in a complementary way and offer domain-specific data services. MobiSpaces paves the way towards future mobility data spaces, in line with European's Union directives, aiming at more efficient, trustworthy and reliable data sharing across organizations and businesses.

Acknowledgments

The research leading to the results presented in this document has received funding from the European Union's funded Project MobiSpaces under grant agreement no 101070279.

References

[1] 2023. *White paper: City data spaces: A guide to building and operationalising data services.* Technical Report. SmartCitiesWorld & FIWARE. 24 pages.
[2] Edward Curry (Ed.). 2020. *Real-time Linked Dataspaces - Enabling Data Ecosystems for Intelligent Systems.* Springer.
[3] Edward Curry, Simon Scerri, and Tuomo Tuikka. 2022. Data Spaces: Design, Deployment and Future Directions. https://doi.org/10.1007/978-3-030-98636-0_1
[4] Tobias Dam, Lukas Daniel Klausner, Sebastian Neumaier, and Torsten Priebe. 2023. A Survey of Dataspace Connector Implementations. In *Proc. of ITADATA (CEUR Workshop Proceedings, Vol. 3606).* CEUR-WS.org.
[5] Christos Doulkeridis et al. 2023. MobiSpaces: An Architecture for Energy-Efficient Data Spaces for Mobility Data. In *IEEE International Conference on Big Data, BigData 2023, Sorrento, Italy, December 15-18, 2023.* IEEE, 1487–1494.
[6] Christos Doulkeridis, Akrivi Vlachou, Nikos Pelekis, and Yannis Theodoridis. 2021. A Survey on Big Data Processing Frameworks for Mobility Analytics. *SIGMOD Rec.* 50, 2 (2021), 18–29.
[7] Michael J. Franklin, Alon Y. Halevy, and David Maier. 2005. From databases to dataspaces: a new abstraction for information management. *SIGMOD Rec.* 34, 4 (2005), 27–33.
[8] GAIA-X. 2022. Gaia-X Architecture Document 22.10. https://docs.gaia-x.eu/technical-committee/architecture-document/22.10/
[9] GAIA-X. 2023. Gaia-X Architecture Document 23.10. https://docs.gaia-x.eu/technical-committee/architecture-document/23.10
[10] Anita Graser, Christos Doulkeridis, and George S. Theodoropoulos. 2023. Green Mobility Data Spaces. *ERCIM News* 2023, 135 (2023).
[11] Anita Graser, Peter Widhalm, and Melitta Dragaschnig. 2020. The M^3 massive movement model: a distributed incrementally updatable solution for big movement data exploration. *International Journal of Geographical Information Science* 34, 12 (2020), 2517–2540.
[12] Marco Hauff, Lina Molinas Comet, Paul Moosmann, Christoph Lange, Ioannis Chrysakis, and Johannes Theissen-Lipp. 2024. FAIRness in Dataspaces: The Role of Semantics for Data Management. In *Proc. of SDS.*
[13] Georgios Makridis, Vasileios Koukos, Georgios Fatouros, and Dimosthenis Kyriazis. 2023. Enhancing Explainability in Mobility Data Science through a combination of methods. *arXiv preprint arXiv:2312.00380* (2023).
[14] Philip Mavrepis, Georgios Makridis, Georgios Fatouros, Vasileios Koukos, Maria Margarita Separdani, and Dimosthenis Kyriazis. 2024. XAI for All: Can Large Language Models Simplify Explainable AI? *arXiv preprint arXiv:2401.13110* (2024).
[15] Mohamed F. Mokbel et al. 2023. Towards Mobility Data Science (Vision Paper). *CoRR* abs/2307.05717 (2023). https://doi.org/10.48550/ARXIV.2307.05717
[16] Boris Otto, Michael Hompel, and Stefan Wrobel. 2019. *International Data Spaces: Reference architecture for the digitization of industries.* 109–128. https://doi.org/10.1007/978-3-662-58134-6_8
[17] Boris Otto, Michael ten Hompel, and Stefan Wrobel. 2022. *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage.* Springer Nature.
[18] Boris Otto, Michael ten Hompel, and Stefan Wrobel. 2022. Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. https://doi.org/10.1007/978-3-030-93975-5
[19] Sebastian Pretzsch, Holger Drees, and Lutz Rittershaus. 2022. Mobility Data Space: A Secure Data Space for the Sovereign and Cross-Platform Utilization of Mobility Data. In *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage.* Springer, 343–361.
[20] Daniela Pöhn, Michael Grabatin, and Wolfgang Hommel. 2021. eID and Self-Sovereign Identity Usage: An Overview. *Electronics* 10, 22 (2021).
[21] Georgios M. Santipantakis, Christos Doulkeridis, and George A. Vouros. 2023. An Ontology for Representing and Querying Semantic Trajectories in the Maritime Domain. In *Proc. of ADBIS*, Vol. 13985. Springer, 224–237.
[22] Georgios M. Santipantakis, Konstantinos I. Kotis, Apostolos Glenis, George A. Vouros, Christos Doulkeridis, and Akrivi Vlachou. 2022. RDF-Gen: generating RDF triples from big data sources. *Knowl. Inf. Syst.* 64, 11 (2022), 2985–3015.
[23] Georgios M. Santipantakis, Konstantinos I. Kotis, George A. Vouros, and Christos Doulkeridis. 2018. RDF-Gen: Generating RDF from Streaming and Archival Data. In *Proc. of WIMS.* ACM, 28:1–28:10.
[24] Raghav Sethi, Martin Traverso, Dain Sundstrom, David Phillips, Wenlei Xie, Yutian Sun, Nezih Yegitbasi, Haozhun Jin, Eric Hwang, Nileema Shingte, and Christopher Berner. 2019. Presto: SQL on Everything. In *Proc. of ICDE.* 1802–1813.
[25] Hubert Tardieu. 2022. *Role of Gaia-X in the European Data Space Ecosystem.* Springer International Publishing, Cham, 41–59.
[26] Andreas Tritsarolis. 2022. Towards understanding privacy-aware artificial intelligence. (2022).
[27] George A. Vouros et al. 2019. The datAcron Ontology for the Specification of Semantic Trajectories - Specification of Semantic Trajectories for Data Transformations Supporting Visual Analytics. *J. Data Semant.* 8, 4 (2019), 235–262.
[28] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3 (2016).